



Infrared spectroscopy in the gas and liquid phase from first principle molecular dynamics simulations - Application to small peptides

Marie-Pierre Gageot, Michael Martinez, Rodolphe Vuilleumier

► To cite this version:

Marie-Pierre Gageot, Michael Martinez, Rodolphe Vuilleumier. Infrared spectroscopy in the gas and liquid phase from first principle molecular dynamics simulations - Application to small peptides. Molecular Physics, 2008, 105 (19-22), pp.2857-2878. 10.1080/00268970701724974 . hal-00513148

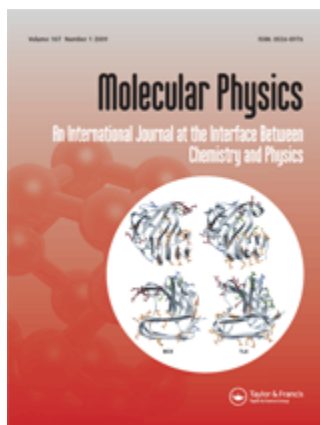
HAL Id: hal-00513148

<https://hal.science/hal-00513148>

Submitted on 1 Sep 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Infrared spectroscopy in the gas and liquid phase from first principle molecular dynamics simulations - Application to small peptides

Journal:	<i>Molecular Physics</i>
Manuscript ID:	TMPH-2007-0165.R1
Manuscript Type:	Full Paper
Date Submitted by the Author:	17-Sep-2007
Complete List of Authors:	Gaigeot, Marie-Pierre; University Evry val d'Essonne, LAMBE UMR 8587 Laboratoire Analyse et Modelisation pour la Biologie et l'Environnement Vuilleumier, Rodolphe; University P/M Curie, LPTMC UMR 7600 Martinez, Michael; University P/M Curie, LPTMC UMR 7600
Keywords:	Car-Parrinello molecular dynamics, spectroscopy, vibration, temperature, peptides
Note: The following files were submitted by the author for peer review, but cannot be converted to PDF. You must view these files (e.g. movies) online.	
figures.tar.gz article.tex	



Infrared spectroscopy in the gas and liquid phase from first principle molecular dynamics simulations - Application to small peptides

Marie-Pierre Gaigeot^{1,2,*}, Michaël Martinez^{2,†} and Rodolphe Vuilleumier²

¹ LAMBE Laboratoire Analyse et Modélisation pour la Biologie et l'Environnement, UMR-CNRS 8587
 Université d'Evry val d'Essonne, Blvd F. Mitterrand, Bât. Maupertuis, F-91025 Evry, France, gaigeot@ccr.jussieu.fr

² LPTMC Laboratoire de Physique Théorique de la Matière Condensée, UMR-CNRS 7600
 Tour 24-25, 2^{ème} étage, c.c. 121
 Université Pierre et Marie Curie, 4 place Jussieu, F-75005 Paris, France, vuilleum@lptmc.jussieu.fr

(Received 00 Month 200x; in final form 00 Month 200x)

We discuss the applicability of finite temperature Car-Parrinello molecular dynamics simulations for the calculation of infrared spectra of complex molecular systems, either in the gas phase or in the condensed phase, taking examples from the infrared spectroscopy of N-methylacetamide and small peptides. Band assignments for the simulation is still challenging and we introduce here a general method for obtaining effective normal modes of molecular systems from Molecular Dynamics simulations. The effective normal modes are defined as linear combination of internal coordinates such that the power spectra of these modes are as localized as possible in frequency. We further define band intensities for these modes from different levels of approximation of the infrared spectrum. Applications of this approach for assigning infrared bands from first-principle molecular dynamics simulations are presented for N-methylacetamide in gas phase and in solution, for the gas phase alanine dipeptide and the gas phase octa-alanine peptide.

1 Introduction

Vibrational spectroscopy (Infrared and Raman) has been recognized for long as a fundamental tool for the characterization and understanding of the chemical state of a molecule, the structure and dynamics of the molecule and its environment. Vibrational spectroscopy is in particular commonly applied for the characterization of secondary structures of peptides and proteins in solution or membranes [1–4]. As an illustration, structural motifs such as helices or beta sheets can be clearly identified from each others by changes in frequency positions, shapes and intensities of the infrared or Raman bands [2, 5, 6].

Linear absorption Infrared spectroscopy has been successfully applied for the last two decades on all kinds of molecular systems in the liquid, solid and gas (matrix) phases. But the past decade has seen the developement of new vibrational spectroscopy methods. Pump-probe non linear 2-dimensional infrared spectroscopy 2D-IR [7–12] is one of these methods. It has been developed in order to circumvent the limitations of linear IR, in particular bringing forth the couplings between modes and the extra-information it gives in the understanding of molecular structures. The advance in 2D-IR is the time dependence recording of the infrared fingerprints: dynamics of a molecular system is thus directly at hand with these experiments, in real time. Up to now, the methodology has been principally developed on the Amide I mode (stretching C=O cabonyl) and mainly applied on biomimetic organic molecules such as peptide models in the liquid phase. Same developments are being made with multi-dimensional Raman spectroscopy [13, 14]. Gas phase spectroscopy has also been revolutionarized during the past decade, going far beyond the low-temperature matrix gas phase spectroscopy. Hence, gas phase mass spectrometry and infrared vibrational spectroscopy have been gathered in a set-up which thus couples Mass Spectrometry (MS) with Infrared Multi Photon Dissociation (IR-MPD), as devised at FELIX [15] or CLIO [16] in Europe. This

*Corresponding authors: gaigeot@ccr.jussieu.fr & vuilleum@lptmc.jussieu.fr

†Present adress: Molecular & Cellular Modeling Group, EML Research gGmbH, 69118 Heidelberg, Germany

is an action spectroscopy, where the fragment ion yields are recorded as a function of the IR excitation energy. Conditions necessary for IR-MPD of gas phase molecules are fulfilled by using an infrared free electron laser (FEL). Thermalization of the produced ions is an important aspect of these experiments. Though the ions are initially produced at high internal energies, their trapping in an ICR (Ion Cyclotron Resonance) cell during a few seconds before exposure to the IR radiation ensures their relaxation and thermalization at room temperature. This device therefore gives the IR spectroscopy of trapped gas phase ions at room temperature, close enough to relevant physiological temperatures of biomolecules. See for instance refs. [17, 18] for more information on the subject.

We are thus able to vibrationally probe molecules either in the gas phase or in the liquid phase, at finite temperature: a direct comparison between the gas phase and liquid phase vibrational patterns can therefore be conducted, *at the same finite temperature*. Environmental effects, such as the aqueous hydrogen bonding to biomolecules, can be directly probed and characterized from this comparison.

Assignment of vibrational modes is an important step in the structural and dynamical interpretation of the spectral fingerprints. Here is where modelling and computation can be most helpful, in particular when a full spectrum, including positions, shapes and intensities, can be computed and directly compared to experimental data. This however necessitates accurate intra- and inter-molecular potentials, able to both describe the typical intramolecular motions of the molecule under study but also the induced shifts due to conformational changes or due to the interaction with the environment. Such potential should then correctly incorporate anharmonic terms and polarisation effects, in particular at finite temperature.

This is in a way exactly the same statement as in the seminal work of Péter Pulay [19] ! Since then, great progress has though been made on the computation and interpretation of vibrational spectra of molecules of reasonable size and complexity, such as bio-organic molecules which are of main interest for many research groups nowadays.

The conventional way of calculating infrared spectra principally involves performing ab initio chemistry calculations [18, 20]. This is the traditional approach for dealing with small organic molecules. In such a scheme, geometry optimizations at 0 K are performed and vibrational analyses in the harmonic approximation are subsequently done for each equilibrium conformation. A match between experimental and calculated IR patterns obtained for the different conformers is generally checked in terms of positions and relative intensities of IR bands. The bottleneck of such theoretical calculations is mainly the determination of all the representative equilibrium conformations of the molecule. This is a tremendous amount of work for flexible and floppy molecules. A small di-peptide already possesses several conformers (representative of those encountered in larger peptides) which are generally lying within a small energy range. At room temperature, one can thus expect conformational isomerisation dynamics to take place. In other words, the peptide does possess enough internal energy to explore different zones of its potential energy surface (PES), particularly if finite temperature IR-MPD gas phase experiments are involved. Therefore, different isomeric conformations of the molecule can be accessible at that temperature and will most likely play a role on the peptide IR spectroscopy. Another limitation (and not the least) of quantum chemistry calculations is the modeling of the environment surrounding the vibrationally probed solute molecule, water being the most ubiquitous solvent. Properly characterising the surrounding hydration shells is indeed out of reach of such calculations, where only a few water molecules that belong to the first hydration shell and are hydrogen bonded to the main hydrophilic sites of the solute can be incorporated into the calculation.

Taking into account the dynamics of the molecule and its consequences on the measured properties can only be achieved through molecular dynamics simulations (MD). MD simulations are also the only approach capable of modeling condensed phase systems, such as homogeneous liquids or a solute immersed in liquid water (or any other solvent environment). Lately, Density Functional Theory (DFT) based Molecular Dynamics has proven successful for the computation of vibrational spectra at finite temperature. This development is the result of the latest advances in density functional theory (DFT) based molecular dynamics simulation ("Car-Parrinello" [21]), brought about by the implementation of the modern theory of polarization and the maximally localized Wannier functions. In Car-Parrinello molecular dynamics simulations the forces acting on the atoms are obtained at each time-step of the simulation by an electronic structure calculation in the DFT framework. This thus combines accurate description of forces in many

environments and situations, while providing a description of the structure and dynamics of a system in its fluctuating environment at finite temperature. Since the pioneering work of Silvestrelli *et al.* for the computation of the infrared spectrum of liquid water [22], DFT-based MD has been applied successfully to the calculation of IR spectra of pure liquids [23,24] or of small biological molecules in aqueous solution [25–29], molecular crystals [30] or Raman spectra in the condensed phase [31–33]. As we have demonstrated in recent studies of the IR absorption of uracil [34], N-Methyl-Acetamide [35], and the protonated alanine peptide [18,36], either in liquid water or in the gas phase, it is now possible to perform a detailed infrared calculation for models of bio-organic molecules in finite temperature bulk solution with all water molecules treated at the same level of theory as the solute, as well as in finite temperature gas phase.

An accurate calculation of infrared spectra is one issue to achieve, the assignment of the active bands into individual atomic displacements or vibrational modes is another one, and this issue is essential to the understanding of the underneath molecular structural and dynamical properties.

We have recently developed a systematic method for decomposing vibrational spectra as an approximate sum of effective normal mode contributions [37]. We have defined these effective normal modes from the data of the Fourier Transform of Velocity Correlation Functions so that they can be extracted directly from the MD trajectory. In this framework, the effective normal modes are linear combinations of atomic displacements constructed such that the corresponding power spectra is as localized as possible in frequency. There is no loss of information from this transformation which is just a change of coordinates.

The present paper aims at demonstrating the applicability of *ab initio* molecular dynamics simulations (such as the Car-Parrinello approach in the DFT framework) for the calculation of infrared spectra of bio-organic molecules (in relation to gas phase and liquid phase finite temperature experiments) and our methodology for the extraction and assignment of vibrational modes in terms of internal coordinates. We will concentrate on the infrared spectroscopy of peptide models, either in the gas phase or in the liquid phase, in relation with finite temperature linear IR or IR-MPD experiments. We will begin with a summary review of static *versus* dynamics calculations of infrared spectra, presenting the advantages and limitations of each method. The localization method we have developed will be described next, as well as a way for the calculation of infrared intensities in our framework. Applications will be presented, mainly on the N-Methyl-Acetamide molecule, Alanine di-peptides and the octa-alanine peptide chain, in the gas phase and in the liquid phase. Ongoing further developments will be mentioned at the end of the paper.

2 Static *versus* dynamic *ab initio* calculations of infrared spectra

The basic definition of an infrared spectrum calculation lies in the Fermi Golden Rule which states that the infrared spectrum is calculated through the following expression [38]:

$$I(\omega) = 3 \sum_i \sum_f \rho_i |\langle f | \mathbf{E} \cdot \mathbf{M} | i \rangle|^2 \delta(\omega_{fi} - \omega) \quad (1)$$

where \mathbf{E} is the applied external field vector, \mathbf{M} is the dipole vector of the molecular system, $|i\rangle$ and $|f\rangle$ are respectively the initial and final vibrational states of the system. These are eigenstates of the system excluding the radiation. ρ_i is the density of molecules in the initial vibrational state $|i\rangle$. ω is the reciprocal wave-number (in cm^{-1}) and ω_{fi} is the reciprocal wave-number associated with the transition between the initial and final vibrational states of the system.

• Infrared spectroscopy with static *ab initio* calculations

The root of vibrational spectra calculations is solving the Schrödinger vibrational equation [39,40]. It is expressed within the harmonic normal mode coordinates q_k ($k = 1 \cdots 3N - 6$, where N is the number

of atoms of the investigated molecular system) in order to be analytically and numerically solvable very easily. With these coordinates, the kinetic energy T and potential energy V that appear in the Schrödinger vibrational equation are simply expressed as:

$$T = \frac{1}{2} \sum_{k=1}^{3N-6} \dot{q}_k^2 \quad V = \frac{1}{2} \sum_{k=1}^{3N-6} \lambda_k q_k^2$$

where λ_k is the frequency associated to the normal mode q_k , and the Schrödinger vibrational equation is:

$$-\frac{\hbar^2}{8\pi^2} \sum_{k=1}^{3N-6} \frac{\partial^2 \Psi_V}{\partial q_k^2} + \frac{1}{2} \sum_{k=1}^{3N-6} \lambda_k q_k^2 \Psi_V = E_V \Psi_V \quad (2)$$

where Ψ_V is the vibrational wavefunction and E_V the corresponding energy; \hbar is the Planck's constant. Due to the normal modes decouplings, the vibrational energy E_V and vibrational wavefunction Ψ_V can be decomposed as

$$E_V = E(q_1) + E(q_2) + \cdots + E(q_{3N-6}) \quad \Psi_V = \Psi(q_1)\Psi(q_2) \cdots \Psi(q_{3N-6})$$

so that equation (2) is satisfied if the following $3N-6$ decoupled equations are satisfied

$$-\frac{\hbar^2}{8\pi^2} \frac{\partial^2 \Psi(q_k)}{\partial q_k^2} + \frac{1}{2} \lambda_k q_k^2 \Psi(q_k) = E_k \Psi(q_k) \quad \forall k = 1 \cdots 3N-6 \quad (3)$$

Solution of the Schrödinger vibrational equation of a harmonic oscillator q_k (equation 3) is well-known, with the vibrational energy $E_k = (v_k + 1/2)\hbar\nu_k$ (where v_k is a quantum number associated with the normal frequency ν_k), and the vibrational wavefunction $\Psi_{v_k}(q_k) = C_k \exp(-1/2\gamma_k q_k^2) H_{v_k}(\gamma_k q_k)$ where C_k is a normalising factor, $\gamma_k = 4\pi^2\nu_k/\hbar$ and $H_{v_k}(\gamma_k q_k)$ is the Hermite polynomial of degree v_k in q_k . The harmonic oscillator wavefunctions are orthonormalised, i.e.:

$$\int dq_k \Psi_{v_k}^*(q_k) \Psi_{v'_k}(q_k) = \delta_{v_k v'_k} \quad (4)$$

The intensity of a spectral line is the probability of the transition which gives rise to the line: the coefficient of absorption of light of frequency $\nu_{v_k v'_k}$ for infrared spectroscopy is given by the Einstein coefficient [41]

$$B_{v_k v'_k} = \frac{8\pi^3}{3\hbar^2} \left[|\mu_{X v_k v'_k}|^2 + |\mu_{Y v_k v'_k}|^2 + |\mu_{Z v_k v'_k}|^2 \right]$$

where μ_X , μ_Y , μ_Z are the components of the transition dipole moment:

$$\mu_{\alpha V V'} = \int d\tau \Psi_V^* \mu_{\alpha} \Psi_V \quad \alpha = X, Y, Z \quad (5)$$

with μ_{α} the α^{th} (X, Y, Z) component of the electric dipole moment of the molecular system. Ψ_V and $\Psi_{V'}$ are the total vibrational wavefunctions for states V and V' . The integral is taken over the configuration space of the system.

By application of the electric harmonic approximation (second harmonic approximation), each μ_α component can be written:

$$\mu_\alpha = \mu_\alpha^{(0)} + \sum_{k=1}^{3N-6} \mu_\alpha^{(k)} q_k$$

where $\mu_\alpha^{(0)}$ is the permanent dipole moment of the molecule evaluated at the equilibrium geometry, and $\mu_\alpha^{(k)} = \left(\frac{\partial \mu_\alpha}{\partial q_k} \right)_0$ is also evaluated at the equilibrium geometry. Equation (5) now reads:

$$\mu_{\alpha VV'} = \mu_\alpha^{(0)} \int d\tau \Psi_{V'}^* \Psi_V + \sum_{k=1}^{3N-6} \left(\frac{\partial \mu_\alpha}{\partial q_k} \right)_0 \int d\tau \Psi_{V'}^* q_k \Psi_V \quad (6)$$

which gives rise to the well-known infrared selection rules of harmonic oscillators (with orthonormalized wave-functions, eq. 4), i.e. $\mu_{\alpha VV'} \neq 0$ if $v_{l'} = v_l$ for all vibrational levels except $v_{k'}$ and v_k for which $\int dq_k \Psi_{v_{k'}}^* q_k \Psi_{v_k}$ is different from zero if $v_{k'} = v_k \pm 1$ (absorption or emission of one photon from two consecutive vibrational levels). The permanent dipole $\mu_\alpha^{(0)}$ does not play any role in the intensity of the spectral lines.

The absolute intensity of an active infrared transition is given by [39]

$$\kappa = \frac{8\pi^3}{3ch} \nu_{VV'} (N_{V'} - N_V) \left[|\mu_{XVV'}|^2 + |\mu_{YVV'}|^2 + |\mu_{ZVV'}|^2 \right]$$

where N_V and $N_{V'}$ are the number of molecules per unit volume respectively in the state V and V' , and c is the speed of light. In the double harmonic approximation, κ is thus proportional to the sum of the transition dipole derivatives with respect to the harmonic normal modes, i.e.

$$\kappa = \sum_{\alpha=x,y,z} \sum_{k=1}^{3N-6} \left| \left(\frac{\partial \mu_\alpha}{\partial q_k} \right)_0 \right|^2 \delta_{v_{k'}, v_k \pm 1}. \quad (7)$$

In the double harmonic approximation, the Fermi Golden Rule thus reduces to the calculation of the normal modes of the molecular system in its equilibrium geometry and to the calculation of the transition dipole derivatives $\left(\frac{\partial \mu_\alpha}{\partial q_k} \right)_0$. Normal modes q_k give the frequency value at which a $v_k = 0 \rightarrow v_{k'} = 1$ transition takes place and the associated atomic movements, while dipole derivatives give the infrared intensities associated to each mode. The investigated infrared spectroscopy is related to the absorption of one photon from $v = 0$ to $v = 1$ harmonic vibrational transitions.

This is achieved with geometry optimization and harmonic frequency calculations performed with quantum chemistry calculations. The DFT (Density Functional Theory) framework has the advantage of allowing the study of molecular systems of tens to hundreds of atoms routinely on most of the computer platforms. The conventional way of proceeding is to perform a systematic search for the different possible molecular conformations of lowest energy on the potential energy surface, followed by harmonic frequency calculations of the different identified conformers and the related transition dipole derivatives. Frequencies are usually scaled [20]: the main purpose of the scaling is to compensate for the double harmonic approximations (potential energy surface and dipole moment expression) as well as for the level of the ab initio calculations. The calculated absorption spectra are finally convoluted with a Gaussian or Lorentzian band profile, adjusted in agreement with experimental conditions.

Obviously, the main drawbacks of this approach are *i*) the search for the minima of lowest energy on the potential energy surface, which can be tricky for floppy molecules, *ii*) the double harmonic approximations,

iii) as well as the inherent 0 K temperature associated to the search of the minima on the potential energy surface. This is in particular a very crude approximation for floppy molecules which can undergo conformational dynamics at finite temperature, as should occur in the infrared experiments we will be comparing our calculations to, which are typically performed at temperatures around 300 K. We will come back on these points later on.

• Infrared spectroscopy with ab initio molecular dynamics simulations

Within Linear Response Theory [38,42], the Fermi Golden Rule formula (equation 1) can be rewritten as the Fourier Transform of the dipole moment time correlation function:

$$I(\omega) = \frac{2\pi\beta\omega^2}{3cV} \int_{-\infty}^{\infty} dt \langle \mathbf{M}(t) \cdot \mathbf{M}(0) \rangle \exp(i\omega t) \quad (8)$$

where $\beta = 1/kT$, c is the speed of light in vacuum, V is the volume. The angular brackets indicate a statistical average of the correlation of the dipole moment \mathbf{M} of the absorbing molecular system. This average is taken in the absence of the applied external field. In this formula, we have taken into account a quantum correction factor (multiplying the classical line shape) of the form $\beta\hbar\omega/(1 - \exp(-\beta\hbar\omega))$, which was shown by us and others to give the most accurate results on calculated IR amplitudes [34,43,44]. For a complete discussion on quantum corrections, we refer the reader to refs. [45,46].

Equation (8) gives the whole infrared spectrum of a molecular system with *one single calculation*, i.e. the band positions, the band intensities and the band shapes, through the fourier transform of a time correlation function. There are no approximations made apart from the hypothesis of linear response theory, i.e. a small perturbation on the absorbing molecular system from the applied electric external field.

This is the standard way used in statistical mechanics for calculating the infrared spectrum of a molecular assembly of molecules (isolated molecule, liquids, solutes in the liquid phase, solids) [38,47–55]. Molecular dynamics simulations are adapted to the calculation of the evolution in time of the dipole moment of the system, and therefore its time correlation. We stress again that the only hypothesis on infrared spectra calculations through equation (8) corresponds to the linear response theory in which the applied external electric field is a small perturbation to the absorbing molecular system. In particular, in this kind of calculation there are no harmonic approximations made, be they on the potential energy surface or on the dipole moment, contrary to the static calculations previously described.

The main advantages of the molecular dynamics approach for the calculation of infrared spectra can be summarized as follows.

i) Dynamics simulations are performed at finite temperature; this will mainly be room temperature in our calculations as they will be compared with experiments typically done at this temperature. At finite temperature, there can be a conformational dynamics between different isomeric conformations of the absorbing molecules of interest. All conformations populated when going from one basin to the others on the potential energy surface are thus taken into account in the calculation of the infrared spectrum. This population dynamics therefore gives rise to a natural broadening of the calculated IR active bands, which is essential for the comparison to the experimental spectra. This has been demonstrated by us in the case of the floppy protonated peptide Ala-Ala-H⁺ in the gas phase [18,36]. A very good agreement of the calculated Car-Parrinello infrared spectrum of this molecule with the IR-MPD (InfraRed Multi Photon Dissociation) experimental spectrum, obtained at a finite temperature of ~300 K, could be achieved because the simulated room-temperature dynamics of the gas phase peptide was able to take into account the continual conformational dynamics between the two major isomers of the molecule. This could not be achieved with standard static ab initio calculations [18,56]. The main thing is that we found that the most populated conformations of Ala-Ala-H⁺ were not the geometries at the bottom wells (identified with static calculations) but rather all conformations explored in going from the basin of one isomer to the basin of the other isomer.

ii) When calculating IR spectra from the dipole time correlation function, all anharmonic effects are naturally described. This is to be opposed to the two successive harmonic approximations usually adopted for the determination of IR spectra from static ab initio calculations (harmonic approximation of the potential energy surface at the optimised geometries and electrical harmonic approximation for the transition dipole moments). Both approximations are released in molecular dynamics, simply because they are not needed. In fact, the finite temperature dynamics takes place on all accessible parts of the potential energy surface, be they harmonic or anharmonic. The quality of the potential energy surface is entirely contained in the "ab-initio" force field used in Car-Parrinello dynamics, calculated at the DFT/BLYP level in our works. The good reproduction of the relative positions of the different active bands in our studies (gas phase [36] and liquid phase [24,34,35]) is a demonstration that this level of theory is correct. Moreover, the calculation of IR spectra with molecular dynamics is related only to the time-dependent dipole moment of the molecule, and it does not require any harmonic expansion of the transition dipole moments. Therefore, if the dipole moments and their fluctuations are accurately calculated along the trajectory, the resulting IR spectrum should be reliable too.

iii) One should also emphasize two more points showing that dynamics (in particular Car-Parrinello dynamics) goes beyond ab initio static calculations applied currently in the domain. The first point is connected with the calculation of the infrared spectrum of complex flexible molecules in the gas phase, for example large peptides and nucleic acids. These complex molecular systems display numerous energetically equivalent conformations which are difficult to be characterised with geometry optimisation searches. Dynamics does release this difficulty as the geometry at the minima on the potential energy surface is not needed. On the other hand, molecular dynamics is the natural method of statistical mechanics for modeling complex molecular systems immersed in liquids. This is obviously very important when considering biomimetic molecules which natural environment is liquid water. The characterisation of the vibrational finger-prints of these molecules should better be performed in the liquid phase if one wants to get a proper structural information to be compared with the natural environment of the molecule. Inclusion of a few water molecules hydrogen bonded to the main hydrophilic sites of the molecule is nowadays commonly performed in ab initio geometry optimisations and harmonic frequency calculations, though still challenging. Properly characterising the surrounding hydration shells is nonetheless out of reach of static calculations. Calculations proposed by some authors (see for instance refs. [57–59]) with up to tens of water molecules clustered around the solute are still not convincing as these clusters do not possess any temperature. Water clusters surrounding a biomimetic solute at 0 K can not mimic the 'true' solvation shell of the solute at finite 300 K temperature. There are in particular hydrogen bonds between the solute and water displayed in solute-water clusters at 0 K that will only exist because there is no temperature in the molecular system, or because the full solvation shell (i.e. two or three surrounding shells) is not taken into account in the calculation. This has been shown by us in refs. [34,59] in the case of the solvated uracil.

3 Generalities on Car-Parrinello molecular dynamics simulations

In ab initio Car-Parrinello molecular dynamics (CPMD) [21], the nuclei are treated classically and the electrons quantum mechanically within the DFT formalism. CPMD simulations consist in solving Newton's equations of motion at finite temperature, with the forces that act on the nuclei deriving from the Kohn-Sham energy. Plane-waves and pseudo-potentials are used to represent valence electrons. Contrary to Born-Oppenheimer MD where the Schrödinger equation for the electronic configuration of the system is solved at each time step of the dynamics (i.e. at each new configuration of the nuclei), the Schrödinger equation is solved only once at the beginning of CPMD simulations. The wave-function is subsequently propagated in time adiabatically with the nuclei propagation. This is denoted "the fictitious dynamics of the electrons", and relies on the use of a fictitious electron mass. Typical fictitious mass values are comprised between 100 and 500 a.u. associated with time steps of 1-5 a.u. A detailed presentation of the

Car-Parrinello molecular dynamics method can be found in refs. [60,61].

The DFT-based Car-Parrinello simulations performed in our work follow the general set-up of *ab initio* molecular dynamics simulations that can be found in our previous publications [34,35,62,63]. We use the Becke, Lee, Yang and Parr (BLYP) gradient-corrected functional [64,65] for the exchange and correlation terms. The one-electron orbitals are expanded in a plane-wave basis set with a kinetic energy cut-off of 70 Ry restricted to the Γ point of the Brillouin zone. Medium soft norm-conserving pseudopotentials of the Martins-Trouillier type [66] are used. The core-valence interaction of C, N and O is treated by *s* and *p* potentials with pseudization radii of 1.23, 1.12 and 1.05 *au*, respectively (taking the same radius for *s* and *p*), while H atoms are treated as a *s* potential with a 0.5 *au* radius. We should mention that the 70 Ry energy cut-off is mostly determined by the convergence of the H and O pseudopotentials. Energy expectations are calculated in reciprocal space using the Kleinman-Bylander transformation [67]. Simulations were performed at constant volume using a fictitious electron mass of 500 *au*, a time step of 5 *au* (0.12 femtoseconds). Liquid simulations have been performed within periodic boundary conditions. Starting configurations have been prepared using classical force field simulations, where we chose the precise number of water molecules in the simulation cell; this is done in order to fix the correct density of such a small simulated liquid system. We refer the reader to Ref. [34,63] for more details on the procedure. Our liquid phase simulations take into account ~ 50 water molecules surrounding the solute molecule (N-methylacetamide presented in the present work), i.e. two complete hydration shells and part of the third one. For gas phase simulations, we refer the reader to ref. [36] for details on the choice of the cubic box length, and the use of the decoupling technique of Martyna and Tuckerman [68] in order to eliminate the effect of the periodic images of the charge density.

The dipole moment of the box cell is calculated with the Berry phase representation, as implemented in the Car-Parrinello framework [69]. Briefly, in the limit where the Γ point approximation applies, the electronic contribution to the cell dipole moment \mathbf{M}_α^{el} (where $\alpha = x, y, z$) is given by [70]:

$$\mathbf{M}_\alpha^{el} = \frac{e}{|\mathbf{G}_\alpha|} \Im \ln z_N \quad (9)$$

where $\Im \ln z_N$ is the imaginary part of the logarithm of the dimensionless complex number $z_N = \langle \Psi | e^{-i\mathbf{G}_\alpha \cdot \hat{\mathbf{R}}} | \Psi \rangle$, \mathbf{G}_α is a reciprocal lattice vector of the simple cubic supercell of length L ($\mathbf{G}_1 = 2\pi/L(1,0,0)$, $\mathbf{G}_2 = 2\pi/L(0,1,0)$, $\mathbf{G}_3 = 2\pi/L(0,0,1)$) and $\hat{\mathbf{R}} = \sum_{i=1}^N \hat{\mathbf{r}}_i$ denotes the collective position operator of the N electrons (or in other words the center of the electronic charge distribution). Ψ is the ground-state wave function. The quantity $\Im \ln z_N$ is the Berry phase, which in terms of a set of occupied Kohn-Sham orbitals $\psi_k(\mathbf{r})$ is computed as $\Im \ln z_N = 2\Im \ln \det \mathbf{S}$ with elements of the matrix \mathbf{S} given by $S_{kl} = \langle \psi_k | e^{-i\mathbf{G}_\alpha \cdot \hat{\mathbf{r}}} | \psi_l \rangle$ [70].

The IR spectra in our work are given as products $\alpha(\omega)n(\omega)$ expressed in cm^{-1} (decadic linear absorption coefficient) as a function of reciprocal wavenumber, ω , in cm^{-1} . The spectra have been smoothed with a window filtering applied in the time domain, i.e. each term of the correlation function $C(t)$ is multiplied by a gaussian function $\exp(-0.5\sigma(t/t_{max})^2)$, where t_{max} is the length of the simulation, and σ is 10 for gas phase simulations and 40 for liquid phase simulations.

When interested in the infrared spectrum of a solute immersed in liquid water, we face additional difficulties, namely separation of the absorption in solute and solvent contributions. Due to the limited statistics, the subtraction method used in experiment is not an option for simulation. The approach, initiated by Parrinello *et al.* [69] and that we first applied in the case of a solute immersed in liquid water in our calculation of the IR spectrum of aqueous uracil [34], is based on a decomposition of charge density using the Maximally localized Wannier Functions scheme of Marzari and Vanderbilt [71]. This enabled us to write the total electronic dipole moment $\mathbf{M} = \mathbf{m}_{solute} + \sum_W \mathbf{m}_W$ as the sum of the solute dipole moment (\mathbf{m}_{solute}) and the sum of the solvent dipole moments (\mathbf{m}_W). As a result, the total dipole correlation function $C_{MM}(t)$ needed in equation (8) ($C_{MM}(t) = \langle \mathbf{M}(t) \cdot \mathbf{M}(0) \rangle$) is resolved into the sum of three contributions $C_{W/W}(t)$, $C_{W/solute}(t)$ and $C_{solute/solute}(t)$ which are respectively the water-water dipole correlation function, the water-solute dipole correlation function and the solute-solute dipole correlation

function. The IR spectrum of the solute is calculated from only the Fourier transform of the self-correlation function $C_{solute/solute}(t)$. As discussed in Ref. [34], one disadvantage of this decomposition is that coupling with the surrounding solvent molecules is taken into account only through intermolecular polarization. The cross term $C_{W/solute}(t)$, which includes the remaining IR contributions arising from correlations between the motion of the solute and the solvent molecules, is ignored. In fact, this term is hard to compute as opposite positive and negative contributions make it difficult to reach convergence within the short simulation times of tens of picoseconds. As a last remark, while fully consistent in the sense that the molecular dipole moments rigorously sum to the cell dipole moment, the use of Wannier function for partitioning the electron density nonetheless introduces a certain arbitrariness for the definition of molecular dipole moments in solution.

An alternative approach to achieve a similar, but not necessarily equivalent, decomposition of the absorption signal is to return to the more fundamental current representation, involving the autocorrelation function of the total current, $\mathbf{j} = \dot{\mathbf{M}} = d\mathbf{M}/dt$:

$$n(\omega)\alpha(\omega) = \frac{2\pi}{3Vk_B T} \int_{-\infty}^{+\infty} \langle \mathbf{j}(0)\mathbf{j}(t) \rangle e^{i\omega t} dt. \quad (10)$$

Note again that in this formula we have taken into account a quantum correction factor of the form $\beta\hbar\omega/(1 - \exp(-\beta\hbar\omega))$. Since in the Born-Oppenheimer approximation the dipole moment of the system depends only on the atomic positions, one can make use of a chain rule to compute the current $\mathbf{j}(t)$ at time t :

$$j^\beta(t) = \sum_{i,\alpha} \frac{\partial M^\beta}{\partial x_i^\alpha}(t) \frac{dx_i^\alpha}{dt}(t) = \sum_{i,\alpha} \frac{\partial M^\beta}{\partial x_i^\alpha}(t) v_i^\alpha(t). \quad (11)$$

where \mathbf{M} is the total dipole moment of the system, \mathbf{x}_i is the position of atom i , and $\frac{\partial M^\beta}{\partial x_i^\alpha}$, $\alpha, \beta = x, y, z$, represent the components of the atomic polar tensor (APT) of atom i . The APT of atom i is thus defined as the derivatives of the dipole moment of the system with respect to infinitesimal displacements of atom i [72, 73]. This tensor is directly related to intensities of infrared bands in the double harmonic approximation [39], as we will show in section 6. Numerous calculations of APT have been performed in the solid phase or in the gas phase [72–75], and recently, a calculation of the APT of water molecules in liquid water has been performed [76].

In our applications in liquid water, the APT tensor was calculated for each atom in the solution using Density Functional Response Theory [73]. This was repeated every 1.2 fs over a ≈ 1 ps interval taken from the full trajectory in solution. In the gas phase calculations at low temperature, only one APT tensor of one conformation has been computed. We have used the Putrino *et al.* [31] implementation of the linear response to an applied electric field in the CPMD code [77]. In this scheme, the macroscopic polarization of the periodically replicated cell is defined using the Berry phase approach of Resta [70, 78]. The APT tensor for the whole system is then obtained from the Maxwell relation:

$$\frac{\partial M^\beta}{\partial x_i^\alpha} = \frac{\partial F_i^\alpha}{\partial \mathcal{E}^\beta} = \frac{\partial^2 E_{tot}}{\partial \mathcal{E}^\beta \partial x_i^\alpha} \quad (12)$$

where \mathcal{E} is an applied uniform electric field and \mathbf{F}_i the force acting on particle i . This allows for the calculation of the $3N$ elements of the system APT tensor through only three linear response calculations ($\beta = x, y, z$) [73].

In order to resolve the infrared spectrum of the whole system into solvent, solute and cross contributions, the total current of the system was separated into a solvent and a solute term:

$$j^\beta(t) = j_{water}^\beta(t) + j_{solute}^\beta(t) \quad (13)$$

by restricting the sum in eq. (11) to atoms of the solvent or the solute only. This natural decomposition of the current then provides a decomposition of the infrared spectrum which is different from the one obtained through the maximally localized Wannier orbitals. The infrared spectrum of the solute is now defined as

$$n(\omega)\alpha_{solute}(\omega) = \frac{2\pi}{3Vk_BT} \int_{-\infty}^{+\infty} \langle \mathbf{j}_{solute}(0) \mathbf{j}_{solute}(t) \rangle e^{i\omega t} dt. \quad (14)$$

A major argument in favor of the current scheme is that any partitioning of the charge density is thus avoided.

In our previous works we have shown that DFT-based Car-Parrinello molecular dynamics simulations (CPMD) yield very accurate infrared spectra of biomimetic molecules in the gas phase or immersed in aqueous solvent at room temperature, in terms of band-positions, band-shapes and band-intensities [34–36]. In particular, we have shown on the prototype gas phase protonated Ala-Ala-H⁺ peptide [36] that CPMD simulations are the proper tool to calculate IR absorption spectra of gas phase molecules undergoing multiple isomeric conformations at room temperature. This methodology has been applied to the IR spectra calculations of molecules such as Uracil [34], N-Methyl-Acetamide [35], and the protonated alanine peptide [36], either in liquid water or in the gas phase. We should mention here that in our applications we have systematically found that our calculated infrared spectra have to be blue-shifted by 100-120 cm⁻¹ so that the *whole* calculated bands can be aligned with their experimental counterparts. This holds true at least in the 800-2000 cm⁻¹ spectral region we have systematically investigated. Therefore, though our CPMD calculations do not give the proper absolute values of band positions, they do give the proper account of band-gaps between the different active bands. Among the effects leading to the very good agreement of our calculated infrared spectra with respect to the experiment (once the global translation has been applied) is the effective inclusion of the anharmonic effects in our finite temperature molecular simulations, unlike static ab initio calculations which are being performed in the harmonic approximation.

We stress again that a *single shift factor or global translation* is applied to the whole spectrum and not a scaling factor. This empirical finding is in contrast to static ab initio calculations where a scaling factor is used to correct the theoretical predictions with respect to the observed frequencies. The origin of this is at the moment unclear to us. Effects of the fictitious mass, which leads to instantaneous Car-Parrinello forces being different from Born-Oppenheimer ones whatever the fictitious mass value [79,80], is certainly important and indeed the blue-shift of 100-120 cm⁻¹ can be reduced when performing Car-Parrinello dynamics with smaller fictitious masses for the propagation of the electronic wave-function. This is though at the cost of more expensive simulations. However, it has been demonstrated in references [79,80] that the Car-Parrinello forces can be brought into good agreement with the BO forces by simply rescaling the ionic masses. Were it so, this would amount once more to a scaling factor and not a global translation. Such rescaling of the atomic masses should then be dependent on the vibrational mode considered, in a way not understood at the moment.

4 Assignment of vibrational modes

An accurate calculation of infrared spectra is one goal to achieve, the assignment of the active bands into individual atomic displacements or vibrational modes is another one, and this issue is essential to the understanding of the underlying molecular structural and dynamical properties. In molecular dynamics simulations, interpretation of the infrared active bands into individual atomic displacements is traditionally done using the vibrational density of states (VDOS) formalism. The VDOS is obtained by Fourier transformation of the atomic velocity auto-correlation functions:

$$VDOS(\omega) = \sum_{i=1,N} \int_{-\infty}^{\infty} \langle \mathbf{v}_i(t) \cdot \mathbf{v}_i(0) \rangle \exp(i\omega t) dt \quad (15)$$

where i runs over all atoms of the investigated system. There are no approximations in this formula. In addition, the VDOS can be decomposed according to each atom type in order to really get an interpretation of the vibrational bands in terms of individual atomic motions. This is done by restraining the sum over i in eq. (15) to the atoms of interest only. The advantage of the VDOS formalism is that all anharmonicities are taken into account in the calculation.

In the case of a harmonic molecular system though, velocities can be written as $\dot{q}_k = \dot{q}_k^0 \sin(\omega_k t + \phi_k)$, where \dot{q}_k^0 is the amplitude taken at the initial time, ω_k the wave-number of the k^{th} normal mode, and ϕ_k the phase, so that equation (15) can be re-expressed as

$$VDOS(\omega) = \sum_{i=1,N} \sum_{k,l} Z_{ik} Z_{il} \int_{-\infty}^{\infty} \langle \dot{q}_k(0) \dot{q}_l(t) \rangle e^{i\omega t} dt$$

where $\dot{\mathbf{x}}_i = \sum_k Z_{ik} \dot{q}_k$ is the transformation between cartesian and harmonic normal mode velocities, and $\langle \dot{q}_k(0) \dot{q}_l(t) \rangle$ is the correlation function between these harmonic velocities. Decoupling of harmonic modes leads to the following expression for the VDOS of a harmonic system:

$$VDOS(\omega) = \sum_{i=1,N} \sum_k Z_{ik}^2 \frac{\langle \dot{q}_k^0 \rangle}{2} \delta(\omega - \omega_k) \quad (16)$$

where $\langle \dot{q}_k^0 \rangle$ is taken as an ensemble average over the initial conditions of the dynamics. The VDOS of each individual atom (restraining the sum $\sum_{i=1,N}$ in equation 16 to one atom at a time) is therefore a sum of dirac functions which are localized at the wave-number of each normal mode ω_k , in the case of a strictly harmonic molecular system. This function is weighted by the amplitude Z_{ik}^2 which quantifies the participation of atom i into the normal mode k . Atoms that do not participate in a certain normal mode k thus make no contribution to the spectral signature of the VDOS.

It is mandatory to note here that VDOS spectra correspond to the vibrational density of states of the investigated molecular system, representing all its vibrational modes. However, only some of these modes will be infrared active or Raman active, so VDOS spectra can by no means substitute for IR or Raman spectra. They are used to interpret and assign vibrational modes which are active in IR or Raman spectra. Finally, when a solute is immersed into a solvent, the decomposition into individual atomic contributions of the VDOS can be interpreted as the contributions of the atoms of the solute on the one hand, and the contributions of the atoms of the solvent on the other hand. It is therefore easy to assess the couplings between solute and solvent atoms and their contributions to the VDOS spectral features of the solute molecule.

This methodology is illustrated in figure 1 in the case of the Car-Parrinello molecular dynamics simulation of N-Methyl-Acetamide in the gas phase at 20 K. We have reported the VDOS of the system as calculated through equation (15) at the bottom of the figure, and the individual VDOS associated with each atom type on top. We only report here the individual contributions of the four peptidic atoms (H, N, C, O) to the VDOS of N-Methyl-Acetamide. This is illustrated in the 1000-2000 cm^{-1} domain where the active bands can still be interpreted by simple combinations of stretching and bending movements of the atoms. Hence, one can note that the four atoms participate to the vibrational band located at $\sim 1600 \text{ cm}^{-1}$, while the oxygen atom does not contribute to the $\sim 1480 \text{ cm}^{-1}$ band. Again, the $\sim 1200 \text{ cm}^{-1}$ band can be decomposed to indicate the contributions of the C, N and H atoms, while the $\sim 1100 \text{ cm}^{-1}$ band displays contributions arising only from N and H atoms. One can thus infer that these bands are respectively related to C=O stretch and N-H bending (1600 cm^{-1}), N-H bending and N-C stretch (1480 cm^{-1}), N-H bending and N-C stretch (1200 cm^{-1}), and N-H bending (1100 cm^{-1}). Nonetheless, the amplitudes of the individual peaks can not be interpreted *per se*, so that we clearly can not conclude on the relative weight of each movement on each vibrational mode. Moreover, contributions arising from the methyl groups have to be considered in order to make a definitive statement on the assignment and interpretation of each

vibrational band. For a complete discussion on N-Methyl-Acetamide, see our ref. [35].

As can be seen from this example, analysis of the vibrational density of states is not so convenient and becomes rapidly fastidious as soon as the number of atoms in the molecular system increases. This analysis becomes basically unreachable for molecules that contain more than 10 atoms, N-Methyl-Acetamide being one of the more complex molecules on which this approach can be easily applied. One main bottleneck of this analysis is that VDOS signatures of each individual atom are spread all over the spectrum, which makes the interpretation and assignment of the vibrational bands relatively difficult. Last, but not least, VDOS do not provide a direct analysis in terms of stretches, bends and torsions, as we need for the interpretation of spectroscopic features of molecules. As we saw before, these motions can be inferred from simple molecules, even of the size of N-Methyl-Acetamide, but this is not true anymore when the size and complexity of the molecule increase. The same argument holds when we assign bands in the far-IR region where movements become more complicated, with couplings of torsional modes. Moreover, a quantitative analysis of the couplings between the movements that participate to the vibrations is out of reach with this approach. Another approach and formalism is required in order to answer to these questions.

Other methods have been proposed in the literature for assigning or approximating vibrational spectra, like spectral analysis [81–84], filtering [85–88], instantaneous normal mode analysis and its variants [89–92], or principal mode analysis (PMA) [93–96] and essential dynamics [97–99], both based on the diagonalisation of covariance matrices. Instantaneous normal mode analysis requires however many calculations of the Hessian system which can become quickly prohibitive for DFT-based MD simulation. Due to its simplicity, PMA has lately attracted attention [94, 95, 100–102].

Our new general definition for effective normal modes, that we will present in the following section, based on a localization principle in frequency space, not only leads at zero temperature to the usual normal mode analysis, but is also consistent with the PMA approach. Moreover, it provides another route to extract an effective Hessian and effective normal modes at finite temperature, without any computational cost.

5 New methodology for the assignment of vibrational modes: localized modes in terms of internal coordinates

For the purpose of assigning bands in vibrational spectra, we propose to construct modes q_k as (invertible) linear combinations of the natural coordinates that describe the dynamics of the systems:

$$\zeta_i(t) = Z_{ik} q_k(t) \Leftrightarrow q_k(t) = Z_{ki}^{-1} \zeta_i(t) \quad (17)$$

$$\dot{\zeta}_i(t) = Z_{ik} \dot{q}_k(t) \Leftrightarrow \dot{q}_k(t) = Z_{ki}^{-1} \dot{\zeta}_i(t). \quad (18)$$

In this equation and in the following, the Einstein summation convention will be used.

The coordinates $\zeta_i(t)$ can be either a set of cartesian coordinates $x_i(t)$ or a set of internal coordinates $S_j(t)$ or can also be mass weighted coordinates. As we saw above, since band assignment from VDOS is limited by the fact that these VDOS are delocalized in frequency we ask here that the mode q_k have VDOS as localized as possible in frequency. Noting $P_k^q(\omega)$ the power spectrum of mode k :

$$P_k^q(\omega) = \int_{-\infty}^{+\infty} \langle \dot{q}_k(0) \dot{q}_k(t) \rangle e^{i\omega t} dt \quad (19)$$

the modes are localized in frequency by minimizing the functional:

$$\Omega^{(n)} = \sum_k \left(\frac{\beta}{2\pi} \int_{-\infty}^{+\infty} d\omega |\omega^{2n}| P_k^q(\omega) - \left(\frac{\beta}{2\pi} \int_{-\infty}^{+\infty} d\omega |\omega^n| P_k^q(\omega) \right)^2 \right) \quad (20)$$

with respect to linear transformation \mathbf{Z} . This functional is parametrized by the free parameter n . In the case $n = 1$, the functional $\Omega^{(1)} = \sum_k (\langle \omega^2 \rangle_k - \langle \omega \rangle_k^2)$ is the sum of the spreads of the power spectra of modes k ; in the following we have rather made the choice $n = 2$.

This minimization criterion however is not sufficient for describing the localized modes q_k as the functional $\Omega^{(n)}$ can be minimized by setting $q_k = 0$. We thus require a normalisation and orthogonality criterion:

$$\frac{1}{2\pi} \int_{-\infty}^{+\infty} d\omega \left(\int_{-\infty}^{+\infty} \langle \dot{q}_k(0) \dot{q}_l(t) \rangle e^{i\omega t} dt \right) = \langle \dot{q}_k(0) \dot{q}_l(0) \rangle = k_B T \delta_{kl} \quad (21)$$

where T is the temperature and k_B the Boltzmann constant. This constraint then imposes the equipartition of the energy in the modes while keeping the modes decorrelated at equilibrium. This criterion will also ensure that the matrix \mathbf{Z} is invertible as required. The choice of introducing the temperature in this equation allows for a definition of the modes independant of T for a truly harmonic system.

To better understand the effect of band localization we can employ the Wiener-Khintchine theorem [103] to show a Cauchy-Schwartz relation:

$$P_{kl}^q(\omega) \leq \sqrt{P_{kk}^q(\omega) P_{ll}^q(\omega)} \quad (22)$$

where $P_{kl}^q(\omega)$ is a generalization of the power spectra (eq. 19) in a matrix form:

$$P_{kl}^q(\omega) = \int_{-\infty}^{+\infty} \langle \dot{q}_k(0) \dot{q}_l(t) \rangle e^{i\omega t} dt \quad (23)$$

If the power spectra $P_{kk}^q(\omega)$ and $P_{ll}^q(\omega)$ are well localized in frequency we can hope that for $k \neq l$ their overlap is very small and as a result the cross-correlation $P_{kl}^q(\omega)$ is very small for all ω such that the modes k and l have very small dynamical correlation. This is indeed expected for effective normal modes describing the dynamics of the system. Furthermore, we have seen in the preceeding section that in an harmonic system the normal mode possess delta like VDOS.

Minimization of the localization functional $\Omega^{(n)}$ does not require an iterative procedure like steepest-descent or conjugate gradients. It can be shown [37] that the solution of the minimization problem is solution of a much simpler linear system. The matrix \mathbf{Z} defining the localized modes from the choice of coordinates ζ_i is then solution of a generalized eigenvalue problem:

$$\mathbf{K}_{\zeta}^{(n)} \mathbf{Z}^{-1T} = \mathbf{K}_{\zeta}^{(0)} \mathbf{Z}^{-1T} \mathbf{\Lambda} \quad \text{with the constraint} \quad \mathbf{Z}^{-1} \mathbf{K}_{\zeta}^{(0)} \mathbf{Z}^{-1T} = \mathbf{I}_{3N} \quad (24)$$

where $\mathbf{\Lambda}$ is a diagonal matrix; the matrices $\mathbf{K}_{\zeta}^{(n)}$ are defined from the power spectra of the original ζ_i coordinates as:

$$\mathbf{K}_{\zeta}^{(n)} = \frac{\beta}{2\pi} \int_{-\infty}^{+\infty} d\omega |\omega|^n \mathbf{P}^{\zeta}(\omega) \quad (25)$$

where the matrix $\mathbf{P}^{\zeta}(\omega)$ is defined similarly as $\mathbf{P}^q(\omega)$ for the coordinates ζ_i and $\beta = \frac{1}{k_B T}$. The eigenvalues $\lambda_k^{(n)}$, diagonal elements of the matrix $\mathbf{\Lambda}$, are equal to:

$$\lambda_k^{(n)} = \int_{-\infty}^{+\infty} d\omega |\omega|^n P_{kk}^q \quad (26)$$

and are thus the averages $\langle \omega^n \rangle_k$ of ω^n over the power spectrum of mode k .

The definition of localized modes from their power spectra then appears to be equivalent to a generalized

eigenvalue problem with the same structure as normal mode analysis at zero temperature. In paper [37] we have employed cartesian coordinates to construct the localized normal modes and we have shown that the case $n = 2$ amounts to a generalization at finite temperature of the usual normal mode analysis where the Hessian at the optimized geometry is replaced by the average Hessian at finite temperature.

However, the use of cartesian coordinates may not be appropriate for large floppy molecules containing mobile groups like rotating CH_3 groups. This can be circumvented by using internal coordinates. To this aim we have taken the coordinates ζ_i as stretches, bends, torsions and out of plane angles. In order to build a complete set of non-redundant internal coordinates we have used the methods devised by Decius [104] or Pulay [19, 105]. Once these modes are defined, we can construct the Wilson **B** matrix that describes the change of internal coordinates with respect to cartesian displacement for each configuration along the dynamics:

$$\mathbf{B} \equiv \frac{\partial S_j}{\partial x_i} \quad (27)$$

This **B** matrix can then be used to evaluate the time-derivatives of the internal coordinates, needed for computing the power spectra matrices, from the cartesian velocities simply by applying

$$\dot{S}_j = \sum_i \frac{\partial S_j}{\partial x_i} \dot{x}_i \quad (28)$$

If we further need to compute accelerations of internal coordinates, numerical time-derivatives of the velocities have been used.

Figures 2 and 3 show the resulting VDOS of the effective normal modes obtained from first-principle molecular dynamics simulations for N-Methyl-Acetamide (NMA) in the gas phase and solvated in liquid water, respectively at low temperature and ambient temperature. For NMA, we have defined a complete set of non-redundant internal coordinates according to Pulay's definitions and localized modes were constructed as linear combinations of these coordinates as described above. In both cases, it can be seen that all modes are well localized in frequency, each showing mainly one peak with a spread usually less than 100 cm^{-1} . In the case of NMA in the gas phase, some modes have a nearly perfect localization indicating their harmonic character. In the liquid phase, overlap between bands can be large, as in the case of C-H stretch bands around 2900 cm^{-1} : these modes are however nearly decoupled from each other, and each mode still exhibits a high level of localization.

The method can be straightforwardly applied to larger systems and figure 4 shows the VDOS of effective normal modes calculated from the Car-Parrinello dynamics of a gas phase alanine dipeptide in the C7_{eq} conformation at 20 K. The first-principle molecular dynamics simulation of this molecule was done using the same general setup as for NMA. The initial set of non-redundant internal coordinates was also constructed following Pulay's prescriptions. At this low temperature, the system is mostly harmonic and the VDOS of the effective normal modes are all very well localized around each normal mode frequencies of this system, with very small or even no overlap between localized modes. This suggests that molecular dynamics simulations at low temperature may be an alternative route for determining normal modes, with the advantage that a perfect geometry optimization is not required contrary to the conventional hessian method. This is a net advantage for systems of increasing size and complexity.

Figure 5 illustrates the effect of the localization of VDOS in frequency in the case of the dynamics of the alanine di-peptide. At the top of the figure, VDOS of the two amide N-H groups of the alanine dipeptide calculated with the usual definition presented in equation (15) restricted to the N and H atoms only of each N-H group (see section 4) are shown, while the bottom of the figure illustrates the VDOS after the localization transformation (see section 5) using non-redundant internal coordinates as defined by Pulay [19]. In this latter case, the localized modes are identified as Amide III modes. Clearly, top of the figure exhibits peaks spread over the $1000\text{-}1800 \text{ cm}^{-1}$ range, in many different regions of the spectrum, and

are thus difficult to use as such to interpret the vibrational spectrum of the molecule. The two effective normal modes describing the two Amide III modes of alanine di-peptide are however perfectly localized around 1100 cm^{-1} . One can moreover note the $\sim 20\text{ cm}^{-1}$ frequency shift between both modes, indicating that the two N-H groups do not exhibit equivalent frequency signatures.

In order to describe the resulting effective normal modes in terms of internal motion, it is necessary to quantitatively determine the contribution of an internal coordinate in an effective normal mode. To this end, we employ a method derived from the Potential Energy Distribution as used in standard normal mode analysis [106–108]. PED is based on the contribution of each internal coordinates to the potential energy of an effective normal mode.

The eigenvectors \mathbf{Z} obtained after diagonalisation of the matrix $\mathbf{K}^{(2)}$ satisfy:

$$\mathbf{Z}^{-1} \mathbf{K}^{(2)} \mathbf{Z}^{-1T} = \mathbf{\Lambda} \Leftrightarrow \sum_{k,l} Z_{ik}^T K_{kl}^2 Z_{lj} = \lambda_{ij} \delta_{ij}$$

from which we deduce the normalization condition:

$$\frac{\sum_{kl} Z_{ki} K_{kl}^2 Z_{li}}{\lambda_{ii}} = 1$$

Then, the matrix P_{ij} that defines the contribution of the internal coordinate i to the normal mode j , can be expressed as [108]:

$$P_{ij} = \frac{\sum_l Z_{ji} K_{jl}^2 Z_{li}}{\lambda_i} \quad (29)$$

We have used here an alternative definition [107] of the PED as a slight variation of Morino's approach. Constructing the matrix $\mathbf{P}(k)$ indexed by the normal mode k as:

$$P_{ij}(k) = \frac{Z_{ik} Z_{jk} K_{ij}^2}{\lambda_k},$$

it has been recognised that the contribution of the internal coordinate i in this effective normal mode can be evaluated in two ways: either as a sum of the i^{th} row (or i^{th} column as the matrix is symmetric) - this is McCarthy's proposal - or as considering only the diagonal terms of $\mathbf{P}(k)$ which are necessarily positive:

$$P_{ii}(k) = P_{ik} = \frac{Z_{ik}^2 K_{ii}^2}{\lambda_k} \quad (30)$$

The resulting contributions need then be normalized and are generally expressed in percentage.

Table 1 shows the PED decomposition of some of the most important modes of NMA in the gas phase at 20 K, in both Cis and Trans conformations, and of Trans-Nma immersed in liquid water at 300 K. In the gas phase, three modes can be clearly identified as the three amide modes with Amide I being a nearly pure C=O stretch. In the liquid phase, this mode appears as a mixing of C=O stretch and N-H bend. Comparison of the gas phase and liquid phase for Trans-NMA shows that the C-N stretch and N-H bend are blue-shifted while the C=O stretch is red-shifted in the liquid phase. The resulting overlap of the C=O stretch and NH bend then leads to Amide modes which are mixed. These shifts are compatible with a picture where the zwitterionic form of the peptide bond is favored in solution, presumably due to the large dielectric constant of water and to hydrogen bonds formed between NMA and the surrounding water molecules. In this zwitterionic form, the CO bond acquires a more important single bond character, less stiff than in the neutral form, while the C-N bond acquires a double bond character. Table 1 also shows

that the other modes localized in this region of the spectrum mostly involve motion of the two terminal methyl groups displaying symmetric and anti-symmetric bends as defined by Pulay's non redundant internal coordinates [19].

Here again, concepts from the normal mode analysis are straightforwardly generalised to effective normal modes extracted from MD simulations and can thus help in the band assignments as well as in comparing results, at finite temperature, in the gas phase and condensed phase.

Another way of describing the calculated effective normal modes is simply by visual inspection of the atomic displacements associated with each mode. These displacements are defined as:

$$\delta x_i(k) = \frac{\partial x_i}{\partial q_k} \quad (31)$$

and can be obtained through the chain rule:

$$\frac{\partial x_i}{\partial q_k} = \sum_j \frac{\partial x_i}{\partial S_j} \frac{\partial S_j}{\partial q_k} \quad (32)$$

in which the matrix $\frac{\partial S_j}{\partial q_k}$ is simply the matrix **Z** defining the effective normal modes. The first term, $\frac{\partial x_i}{\partial S_j}$ is obtained from Wilson's **B** matrix defined as:

$$\mathbf{B} \equiv \frac{\partial S_j}{\partial x_i}. \quad (33)$$

This matrix however is $(3N - 6) \times 3N$ and is thus not invertible to obtain the needed $\frac{\partial x_i}{\partial S_j}$ terms. This can however be achieved by further requiring that the displacements $\delta x_i(k)$ satisfy the six Eckart conditions, i.e. no net translation nor rotation of the investigated molecular system. There also, we follow the general prescription of normal mode analysis in internal coordinates from Crawford and Fletcher [109] which results in:

$$\frac{\partial x_i}{\partial S_j} \equiv \mathbf{B}^- = \mathbf{M}^{-1} \mathbf{B}^T (\mathbf{B}^T \mathbf{M}^{-1} \mathbf{B})^{-1} \quad (34)$$

where **B** is the Wilson's matrix and **M** is the matrix of atomic masses. It is then possible to express the atomic displacement vector corresponding to mode *k* as:

$$\delta x_i(k) = (\mathbf{B}^- \mathbf{Z})_{ik} \quad (35)$$

Figure 6 displays some of the effective normal modes of NMA solvated in liquid water. On these figures, as on the PED reported in table 1, one can identify the participation of C=O stretch, NH bend and C-N stretch to the three Amide modes, while the two other modes displayed whose frequencies are in the same range as the Amide bands are deformations of the terminal methyl groups. See table 1 for the precise assignment of the modes.

We are thus able to extract effective normal modes from molecular dynamics simulations performed at finite temperature, either in the gas phase or for a solute immersed in the condensed phase. These modes are effective in the sense that they include temperature, all anharmonicities of the dynamics (potential energy surface anharmonicities, dipole anharmonicities), and dynamical isomerisation along the dynamics. From the effective normal mode localization, a PED analysis can be applied. These two steps are strictly similar to the analyses conducted in static ab initio calculations, with the net advantage of an underneath molecular dynamics simulation.

6 Infrared intensities

The localization of modes in frequency is very helpful for assigning observed bands in IR spectra. However, not all the computed modes may be observed in the spectra because their IR intensity is low. Sometimes many bands may be present in the same frequency domain and only a few may have significant IR intensity to be observable. Thus for the purpose of assigning IR spectra bands, it is of great interest to also have at disposal the IR intensity associated with each effective normal mode motion. This is the purpose of this paragraph.

Starting from the expression of the infrared spectrum as the Fourier-transform of the autocorrelation function of the dipole or its time derivative (current), see eq. (11), we can re-write the current in the newly introduced effective normal modes q_k as:

$$j^\beta(t) = \sum_k \frac{\partial M^\beta}{\partial q_k}(t) \frac{dq_k}{dt}(t) = \sum_k \frac{\partial M^\beta}{\partial q_k}(t) \dot{q}_k(t) \quad (36)$$

to re-express the infrared absorption as:

$$IR(\omega) = n(\omega)\alpha(\omega) = \frac{2\pi}{3Vk_B T} \int_{-\infty}^{+\infty} \sum_{k,l} \left\langle \left(\frac{\partial M^\beta}{\partial q_k} \dot{q}_k \right) (0) \left(\frac{\partial M^\beta}{\partial q_l} \dot{q}_l \right) (t) \right\rangle e^{i\omega t} dt \quad (37)$$

In this equation the velocity \dot{q}_k of the k^{th} effective normal mode is obtained as:

$$\dot{q}_k(t) = \frac{\partial q_k}{\partial S_t} \frac{\partial S_t}{\partial x_i} \dot{x}_i \quad (38)$$

from the internal coordinates definition. The APT tensors expressed in effective normal mode coordinates is a vector $\frac{\partial M^\beta}{\partial q_k}$ analogous to a time-dependent transition dipole vector. It can be calculated in a similar way using the chain rule:

$$\frac{\partial \mathbf{M}}{\partial q_k} = \frac{\partial \mathbf{M}}{\partial x'_i} \frac{\partial x'_i}{\partial S_t} \frac{\partial S_t}{\partial q_k} \quad (39)$$

where the first term, $\frac{\partial \mathbf{M}}{\partial x'_i}$ is the APT tensor in the laboratory frame, the second term is the generalized inverse of Wilson's matrix B which we have already introduced, while the third term is the matrix transform between effective normal modes and the internal coordinates. The first two terms are time-dependent and need to be calculated on the fly along the dynamics with an inversion of the B matrix for each configuration.

The first approximation to this otherwise exact formula for the infrared spectrum is to assume that the effective normal modes are decorrelated from each others, which gives:

$$IR(\omega) = n(\omega)\alpha(\omega) \approx \frac{2\pi}{3Vk_B T} \int_{-\infty}^{+\infty} \sum_k \left\langle \left(\frac{\partial M^\beta}{\partial q_k} \dot{q}_k \right) (0) \left(\frac{\partial M^\beta}{\partial q_k} \dot{q}_k \right) (t) \right\rangle e^{i\omega t} dt \quad (40)$$

as suggested by the definition of the effective normal modes that should minimize the overlap between their respective power spectra. In this approximation, the infrared spectrum is decomposed as a sum of effective normal modes contributions that can be computed in order to interpret the spectrum and assign observed IR bands. The rest of the paragraph is devoted to further approximations that allow to assign an intensity to each effective normal mode.

The next approximation to this sum of modes is to assume a kind of rotation-vibration decorrelation

assuming that the terms $\frac{\partial M^\beta}{\partial q_k}$ are decorrelated from the mode velocities, so that:

$$\left\langle \left(\frac{\partial M^\beta}{\partial q_k} \dot{q}_k \right) (0) \left(\frac{\partial M^\beta}{\partial q_k} \dot{q}_k \right) (t) \right\rangle = \left\langle \left(\frac{\partial M^\beta}{\partial q_k} \right) (0) \left(\frac{\partial M^\beta}{\partial q_k} \right) (t) \right\rangle \langle \dot{q}_k(0) \dot{q}_k(t) \rangle \quad (41)$$

and

$$IR(\omega) = n(\omega) \alpha(\omega) \approx \frac{2\pi}{3Vk_B T} \sum_k \int_{-\infty}^{+\infty} \left\langle \left(\frac{\partial M^\beta}{\partial q_k} \right) (0) \left(\frac{\partial M^\beta}{\partial q_k} \right) (t) \right\rangle \langle \dot{q}_k(0) \dot{q}_k(t) \rangle e^{i\omega t} dt \quad (42)$$

Noting that $\partial \mathbf{M} / \partial q_k$ is a vector which we can write as a norm times a unit vector: $\partial \mathbf{M} / \partial q_k = |\partial \mathbf{M} / \partial q_k| \mathbf{u}(t)$, and further assuming decorrelation between the two, we can write:

$$IR(\omega) = \frac{2\pi}{3Vk_B T} \sum_k \int_{-\infty}^{+\infty} \left\langle \left| \frac{\partial \mathbf{M}}{\partial q_k} (0) \right| \cdot \left| \frac{\partial \mathbf{M}}{\partial q_k} (t) \right| \right\rangle \langle \mathbf{u}_k(0) \cdot \mathbf{u}_k(t) \rangle \langle \dot{q}_k(0) \dot{q}_k(t) \rangle e^{i\omega t} dt \quad (43)$$

which is our expression for the IR spectrum including mode decoupling and complete rotation-vibration decorrelation.

In a final approximation, we can assume that the term $\langle \dot{q}_k(0) \dot{q}_k(t) \rangle$ describing the modes dynamics varies much more rapidly than the two other terms, which thus can be evaluated at time $t = 0$ only. The infrared spectrum is therefore simply written as:

$$IR(\omega) = \frac{2\pi}{3Vk_B T} \sum_k A_k \int_{-\infty}^{+\infty} \langle \dot{q}_k(0) \dot{q}_k(t) \rangle e^{i\omega t} dt \quad (44)$$

where the A_k 's are the mode intensities:

$$A_k = \left\langle \left| \frac{\partial \mathbf{M}}{\partial q_k} (0) \right|^2 \right\rangle. \quad (45)$$

In this form, the infrared spectrum is a weighted sum of the effective normal modes power spectra, whose intensities are obtained from the Atomic Polar Tensor or transition dipole vectors in a similar way as in the double harmonic approximation. In the examples below, the intensities A_k are obtained using the APT calculated on one single configuration only of the system instead of a true configuration average $\langle \left| \frac{\partial \mathbf{M}}{\partial q_k} (0) \right|^2 \rangle$, this allows for a rapid estimation of band intensities for the purpose of assigning bands.

This is however not necessary in our framework, as upon the change of coordinates made to introduce the effective normal modes, all information about the system dynamics is still retained at this stage and the IR spectrum can still be calculated exactly as in eq. (37) or with a minimal approximation, eq. (40) that still allows for a decomposition of the total spectrum in a sum of effective normal modes contributions.

In the following, we will be presenting calculations of IR intensities following equation (42) with the APT tensor and correlation calculated over one pico-second of the dynamics, or using equation (44) in which only one APT tensor is calculated for the first configuration of the dynamics.

6.1 Applications

Comparisons between the infrared spectrum calculated through the Fourier transform of the dipole correlation function (equation (8)) and the Fourier transform of the localized normal modes (equation (44)) making use of APT tensors is presented in figures 7 and 8. Comparisons are presented for the Trans-NMA molecule in the gas phase (Car-Parrinello dynamics at 20 K) and immersed in liquid water (Car-Parrinello molecular dynamics simulation at 300 K). The whole spectrum is presented in the gas phase, while only

the 1000-2000 cm^{-1} domain is presented in the liquid phase. In both cases, it can be seen that the infrared spectrum of Trans-NMA calculated through equation (44) captures the main features of the exact infrared spectrum calculated through equation (8). Moreover, the result is not sensitive to the choice of the APT tensor as can be concluded from the condensed phase calculation in figure 8, be it sampled over part of the dynamics (here one pico-second) or fixed from the first configuration of the dynamics. In particular, structural deformations taken into account in the APT tensors calculated over one pico-second of the trajectory do not give substantial further informations to the infrared intensities and band widths. In the gas phase, both spectra are similar, with mainly an underestimation of the intensity of the Amide I band at $\sim 1600 \text{ cm}^{-1}$ and of the C-H stretching band above 2000 cm^{-1} when equation (44) is applied. The N-H stretching band above 2000 cm^{-1} is overestimated. In the liquid phase calculation, one can see that the three main Amide bands are present in all calculations, though the Amide II band located at $\sim 1400 \text{ cm}^{-1}$ is merged with the Amide I band in the calculations done through the APT tensors. The disappearance of the Amide II band in the APT calculation of the IR spectrum could be due to the choice of definition of the molecular current from the APT tensor, which gives a completely different decomposition of the density in comparison to the Wannier orbital decomposition. Note here (see our paper [35]) that the Car-Parrinello spectrum, calculated with the Wannier orbital decomposition, gives an IR spectrum of NMA immersed in water which is in very good agreement with the experiment.

Figure 9 reports the 1000-1800 cm^{-1} frequency domain of the calculated infrared spectrum of Trans-NMA obtained from our Car-Parrinello trajectory at 20 K (dashed black line) and its decomposition into its fundamental normal mode contributions as described with our formalism of the IR intensities. The intensities of the bands are reported in table 1. The band located at $\sim 1595 \text{ cm}^{-1}$ clearly comes solely from the Amide I mode (stretching C=O as from 95%, red line). The broad band roughly located between 1300 and 1500 cm^{-1} can be nicely decomposed into three contributions. The 1440 cm^{-1} shoulder on the upper frequency part is due to the Amide II mode (bending N-H as from 42%, N-Terminal methyl symmetric stretching as from 17% and stretching C-N as from 10%, blue line); two modes take part in the central part of the broad band, respectively coming from the C-Terminal asymmetric methyl C-H bendings (85%, purple line) and the N-Terminal symmetric methyl C-H bendings (77%, black line); the lower frequency part is due to the C-Terminal symmetric methyl C-H bendings (95%, orange line). Features arising from the symmetric and asymmetric C-H bendings can thus be nicely separated over the whole infrared spectrum, as well as the contributions arising from each of the terminal methyl groups. Finally, the $\sim 1180 \text{ cm}^{-1}$ band results from the Amide III mode, which is a composition of C-N stretching (32%), N-H bending (30%) and C=O stretching (12%) in our calculation.

The same 1000-1800 cm^{-1} frequency domain decomposition is presented in figure 10 in the case of the calculated infrared spectrum of Trans-NMA from our Car-Parrinello molecular dynamics trajectory in aqueous water at 300 K (black dashed line). The intensities of the bands are reported in table 1 and the transition dipole moments are shown as red arrows in figure 6. As can be easily seen, the $1400\text{-}1600 \text{ cm}^{-1}$ active band is composed of the Amide I mode (now interpreted as 54% of C=O stretching and 38% of N-H bending) on its upper frequency part and of the Amide II mode (interpreted as 48% N-H bending, 24% C-N stretching and 13% N-Terminal C-H symmetric stretching) on its lower part. Both modes thus merge into a broad band in the liquid phase, while both modes are now composed of more mixing between C=O and N-H motions. The active band located at $\sim 1230 \text{ cm}^{-1}$ is only composed of the Amide III mode (interpreted as 38% stretching C-N, 35% bending N-H and 12% stretching C=O, which is very similar to the gas phase interpretation). The band located at $\sim 1350 \text{ cm}^{-1}$ is a merging of the methyl symmetric C-H bending motions, the C-Terminal methyl being at the upper frequency part and the N-Terminal methyl at the lower frequency part.

Another illustration is presented in figure 11 in the case of the gas phase Car-Parrinello trajectory of the octa-alanine peptide chain at $\sim 20 \text{ K}$. As can be seen again (top of figure 11), the approximated infrared spectrum (calculated through equation (42)) is very close to the exact infrared spectrum (calculated through equation (8)), the biggest discrepancy arising again from the intensity of the C-H stretching modes at $\sim 2800 \text{ cm}^{-1}$. The three main bands obtained in the 1000-1800 cm^{-1} frequency domain can be nicely decomposed into individual contributions as proposed at the bottom of figure 11. The band at $\sim 1570 \text{ cm}^{-1}$ is composed of the Amide I C=O stretching modes, arising from the different carbonyl

groups along the octa-peptide chain, the band at $\sim 1420\text{ cm}^{-1}$ results from the Amide II N-H bending modes of the different amide groups along the octa-peptide chain, while the $\sim 1130\text{ cm}^{-1}$ band comes from the Amide III modes. For each Amide band, one can see that each decomposed sub-band is peaked at a different absolute position value, the sum of which giving rise to the broadening of the resulting final band.

7 Discussion–Conclusions

We have reviewed here Car-Parrinello molecular dynamics simulations for the calculation of infrared spectra of complex molecular systems, either in the gas phase or in the condensed phase, like here immersed in a liquid environment, at finite temperature. Although first-principle simulations have proved powerful for this aim, interpretation of the resulting vibrational spectrum is often not straightforward, based on the study of VDOS for selected coordinates as illustrated here in the case of N-Methyl-Acetamide. We have here presented a new formalism in order to extract localized effective normal modes from the trajectories in order to interpret and assign the active vibrational bands of the IR spectra. This method is quite general and can be applied in gas, solid or liquid phase. The effective normal modes thus constructed can be written as linear combinations of linear coordinates, being more adapted to solid state or stiff molecules [37], and we have introduced here effective normal modes as linear combinations of internal coordinates. This is better adapted to situations where part of the system/molecule is floppy.

Our methodology has been illustrated on the N-Methyl-Acetamide (NMA) molecule, on the alanine dipeptide molecule and on the more complex octa-alanine peptide chain. The effective normal modes were constructed here in terms of individual internal coordinates (taken as non-redundant as defined by Pulay [19]). We have shown that our formalism is able to localize effective vibrational modes in the frequency domain, allowing for a clear separation of the different modes.

The construction of the effective normal modes is formally only a change of variables and no information is lost from the dynamics. Only the representation of the dynamics has changed, allowing for an easier band assignment since, in this representation, the VDOS are as localized as possible; they still have however a width and possibly complex shape reflecting the complex dynamics in the system. The modes thus extracted are effective in the sense that they include temperature, all anharmonicities of the dynamics (potential energy surface anharmonicities and dipole anharmonicities), and a complex dynamics such as dynamical isomerisation along the time. We have previously shown [37] that this method is close to PMA for a choice $n = -2$ and cartesian coordinates, and is a generalization at finite temperature of the usual normal mode analysis for the choice $n = 2$. Here, we have also shown that concepts like PED can be straightforwardly adapted to our approach. We are then able to compute PED contributions in order to give a proper and precise assignment of each effective mode as illustrated for the gas phase and solvated NMA.

In this paper, we also introduced infrared intensities calculated for the effective normal modes. These are very helpful indicators for the assignment of bands in the observed or calculated IR spectra. Different ways of calculating the IR intensities associated with each effective mode have been devised and we have shown that they correspond in each case to an approximation of the IR 'exact' spectrum. In the simplest case, it amounts to computing the APT tensor in the new set of effective normal mode coordinates, defining a transition moment similar to the normal mode analysis case first introduced by Pulay [19]. As can be seen in the illustrations on the peptide models, the reconstructed IR spectra hence computed from the band intensity and band VDOS compare very well to the exact ones calculated through the Fourier Transform of the time correlation of the dipole moment.

These intensities, although representing the IR spectra in a simplified way as described above, can be also useful in providing an approximation to the IR spectrum when a full calculation may be difficult or costly (calculation of molecular dipoles by orbital rotation or computation of currents). This should be particularly true for Raman spectra for which the approach presented here can bring a useful approximation at a minimal numerical cost.

The effective normal mode method discussed here can be seen as fitting the best harmonic model to the computed dynamics, as we have previously discussed in cartesian coordinates [37]. This suggests that

the method could be used to reparameterize intramolecular classical force-fields to better represent the molecular vibrations in a given environment, including anharmonicity in an effective way. It is however also expected that mode localization will not be very efficient in strong anharmonic situations, as in the case of a complex molecular dynamics where isomerisation between different isomeric conformations occur. In such situations, the methodology discussed here can be slightly modified in order to extract the influence of each isomeric conformation on the effective normal modes computed from the entire dynamics. The velocity correlations can indeed be weighted by the amount of time spent by the molecule in each of the isomeric basins, and the localization principle can still be applied for vibrations in each basin. This is current work in progress to assess the validity of this approach. Another envisioned development for anharmonicity is the inclusion of higher order terms in the definition of effective modes (e.g. a second order polynomial).

Finally, one of the biggest advantages of first-principle molecular dynamics for the computation of vibrational spectra is the explicit description of the electronic system. This allows, in principle, the prediction of other type of vibrational spectroscopy than infrared spectra, like Raman spectroscopy or vibrational circular dichroism.

Acknowledgments. MPG acknowledges support from Genopole-France through the program 'ATIGE' Action Thématique Incitative de Génopole. The authors thank CINES (Montpellier, France) and IDRIS (Orsay, France) for generous access to their computational facilities.

References

- [1] S. Krimm and J. Bandekar. *Adv. Prot. Chem.*, 38:181, 1986.
- [2] H. Torii and M. Tasumi. *J. Chem. Phys.*, 96:3379, 1992.
- [3] *Biological Applications of Raman Spectroscopy, Vol I*, edited by T. Spiro, Wiley-Interscience, New-York, 1987.
- [4] *Infrared Spectroscopy of Biomolecules*, edited by H. Mantsch and D. Chapman, Wiley-Liss, New-York, 1996.
- [5] R. Huang, J. Kubelka, W. Barber-Armstrong, R. Silva, S.M. Decatur, and T.A. Keiderling. *J. Am. Chem. Soc.*, 126:2346, 2004.
- [6] P. Bour and T.A. Keiderling. *J. Phys. Chem. B.*, 109:5348, 2005.
- [7] J.R. Zheng JR and M.D. Fayer. *J. Am. Chem. Soc.*, 129:4328, 2007.
- [8] I.J. Finkelstein, J.R. Zheng, H. Ishikawa, S. Kim, K. Kwak, and M.D. Fayer. *Phys.Chem.Chem.Phys.*, 9:1533, 2007.
- [9] S. Bagchi S, Y.S. Kim, A.K. Charnley, A.B. Smith, and R.M. Hochstrasser. *J. Phys. Chem. B.*, 111:3010, 2007.
- [10] S. Gnanakaran and R.M. Hochstrasser. *J. Am. Chem. Soc.*, 123:12886, 2001.
- [11] P. Hamm, M. Lim, and R.M. Hochstrasser. *J. Phys. Chem. B.*, 102:6123, 1998.
- [12] S. Woutersen and P. Hamm. *J. Phys. Chem. B.*, 104:11316, 2000.
- [13] P.C. Chen and C.C. Joyner. *J. Mol. Struct.*, 799:23, 2006.
- [14] P.C. Chen and C.C. Joyner. *J. Phys. Chem. A.*, 110:7989, 2006.
- [15] J. Oomens, G. Meijer, and G.V. Helden. *J. Phys. Chem. A.*, 105:8302, 2001.
- [16] J. Lemaire, P. Boissel, M. Heninger, G. Maucclair, G. Bellec, H. Mestdag, A. Simon, S. Le Caer, J.M. Ortega, F. Glotin, and P. Maitre. *Phys. Rev. Lett.*, 89:273002-1, 2002. and references therein.
- [17] J. Oomens, B.G. Sartakov, G. Meijer, and G. Von Helden. *Int. J. Mass Spectrom.*, 254:1, 2006.
- [18] G. Grégoire, M.P. Gaigeot, D.C. Marinica, J. Lemaire, J.P. Schermann, and C. Desfrancois. *Phys. Chem. Chem. Phys.* Accepted for publication in *Phys. Chem. Chem. Phys.*, DOI: 10.1039/b618094a.
- [19] P. Pulay, G. Fogarasi, F. Pang, and J.E. Boggs. *J. Am.Chem.Soc.*, 101:2550, 1979.
- [20] K.K. Irikura, R.D. Johnson, and R.N. Kacker. *J. Phys. Chem. A.*, 109:8430, 2005.
- [21] R. Car and M. Parrinello. *Phys. Rev. Lett.*, 55:2471, 1985.
- [22] P.L. Silvestrelli and M. Parrinello. *J. Chem. Phys.*, 111:3572, 1999.
- [23] Vincent Dubois, Paolo Umari, and Alfredo Pasquarello. *Chem. Phys. Lett.*, 390(1-3):193-198, 2004.
- [24] Jan-Willem Handgraaf, Evert Jan Meijer, and Marie-Pierre Gaigeot. *J. Chem. Phys.*, 121(20):10111-10119, 2004.
- [25] M.P. Gaigeot, R. Vuilleumier, M. Sprik, and D. Borgis. *J. Chem. Theory Comput.*, 1:772, 2005.
- [26] M.-P. Gaigeot and M. Sprik. *J. Phys. Chem. B*, 108:7458, 2004.
- [27] M.-P. Gaigeot and M. Sprik. *J. Phys. Chem. B*, 107:10344, 2003.
- [28] Radu Iftimie and Mark E. Tuckerman. *J. Chem. Phys.*, 122(21):214508, 2005.
- [29] M. Nonella, G. Mathias, and P. Tavan. *J. Phys. Chem. A*, 107:8638, 2003.
- [30] Aneta Jezierska, Jaroslaw J. Panek, Aleksander Koll, and Janez Mavri. *J. Chem. Phys.*, 126:205101, 2007.
- [31] Anna Putrino, Daniel Sebastiani, and Michele Parrinello. *J. Chem. Phys.*, 113:7102, 2000.
- [32] A. Putrino and M. Parrinello. *Phys. Rev. Lett.*, 88:176401, 2002.
- [33] A. Miani, S. Raugai, P. Carloni, and M.S. Helfand. *J. Phys. Chem. B.*, 111:2621, 2007.
- [34] M.-P. Gaigeot and M. Sprik. *J. Phys. Chem. B.*, 107:10344, 2003.
- [35] M.-P. Gaigeot, R. Vuilleumier, M. Sprik, and D. Borgis. *J. Chem. Theor. Comput.*, 1:772, 2005.
- [36] C. Marinica, G. Grégoire, C. Desfrancois, J.P. Schermann, D. Borgis, and M.-P. Gaigeot. *J. Phys. Chem. A.*, 110:8802, 2006.
- [37] M. Martinez, M.P. Gaigeot, D. Borgis, and R. Vuilleumier. *J. Chem. Phys.*, 125:144106, 2006.
- [38] D.A. McQuarrie. *Statistical Mechanics*. Harper-Collins Publishers: New York, 1976.
- [39] E.B. Wilson, J.C. Decius, and P.C. Cross. *Molecular Vibrations*. McGraw-Hill, New-York, 1955.
- [40] G. Zerbi. Introduction to the theory of vibrational frequencies and vibrational intensities. In *Vibrational intensities in infrared and Raman spectroscopy*, by W. B. Person and G. Zerbi. Elsevier Scientific Publishing Company, Amsterdam - Oxford - New York, 1982.
- [41] L. Pauling and E.B. Wilson. *Introduction to quantum mechanics, with applications to chemistry*. Dover, 1985.
- [42] R. Kubo, M. Toda, and N. Hashitsume. *Statistical Physics*, volume II. Springer Verlag, second edition, 1991.
- [43] R. Iftimie and M.E. Tuckerman. Decomposing total IR spectra of aqueous systems into solute and solvent contributions: A computational approach using maximally localized Wannier orbitals. *J. Chem. Phys.*, 122:214508, 2005.
- [44] H. Ahlborn, B. Space, and P.B. Moore. *J. Chem. Phys.*, 112:8083, 2000.
- [45] J. Borysow, M. Moraldi, and L. Frommhold. *Mol. Phys.*, 56:913, 1985.
- [46] R. Ramirez, T. Lopez-Ciudad, P. Kumar, and D. Marx. *J. Chem. Phys.*, 121:3973, 2004.
- [47] D. Levesque, J.J. Weis, P. Marteau, J. Obriot, and F. Fondere. *Mol. Phys.*, 54:1161, 1985.
- [48] Y. Guissani, D. Levesque, J.J. Weis, and D.W. Oxtoby. *J. Chem. Phys.*, 77:2153, 1982.
- [49] T. Tassaing, Y. Danten, M. Besnard, E. Zoidis, J. Yarwood, Y. Guissani, and B. Guillot. *Mol. Phys.*, 84:769, 1995.
- [50] B. Guillot. *J. Chem. Phys.*, 95:1543, 1991.
- [51] R. Vuilleumier and D. Borgis. *J. Chem. Phys.*, 111:4251, 1999.
- [52] A. Staib and D. Borgis. *Chem. Phys. Lett.*, 271:232, 1997.
- [53] W.J. Glover and P.A. Madden. *J. Chem. Phys.*, 121:7293, 2004.
- [54] E.A. Pavlatou, P.A. Madden, and M. Wilson. *J. Chem. Phys.*, 107:10446, 1997.
- [55] P.A. Madden and R.W. Impey. *Chem. Phys. Lett.*, 123:502, 1986.
- [56] B. Lucas, G. Grégoire, J. Lemaire, P. Maitre, J.M. Ortega, A. Rupenyan, B. Reimann, J.P. Schermann, and C. Desfrancois. *Phys.Chem.Chem.Phys.*, 6:2659, 2004.
- [57] V.I. Danilov, T. van Mourik, and V.I. Poltev. *Chem. Phys. Lett.*, 429:255, 2006.
- [58] O.V. Shishkin, L. Gorb, and J. Leszczynski. *J. Phys. Chem. B.*, 104:5357, 2000.
- [59] M.P. Gaigeot and M. Ghomi. *J. Phys. Chem. B.*, 105:5007, 2001.
- [60] D. Marx, M. E. Tuckerman, and M. Parrinello. *J. Phys.: Condens. Matter A*, 12:153-159, 2000.
- [61] R. Vuilleumier. Density functional theory based ab initio molecular dynamics using the car-parrinello approach. In *Computer Simulations in Condensed Matter Systems: From Materials to Chemical Biology Volume 1*, by Mauro Ferrario, Kurt Binder, and Giovanni Ciccotti. Springer, Berlin, 2006.
- [62] M.-P. Gaigeot and M. Sprik. *J. Phys. Chem. B.*, 108:7458, 2004.
- [63] R. Vuilleumier and M. Sprik. *J. Chem. Phys.*, 115:3454, 2001.
- [64] A. Becke. *Phys. Rev. A.*, 38:3098, 1988.
- [65] C. Lee, W. Yang, and R.G. Parr. *Phys. Rev. B.*, 37:785, 1988.
- [66] N. Trouillier and J.L. Martins. *Phys. Rev. B.*, 43:1993, 1991.

- [67] L. Kleinman and D.M. Bylander. *Phys. Rev. Lett.*, 48:1425, 1982.
- [68] G.J. Martyna and M.E. Tuckerman. *J. Chem. Phys.*, 110:2810, 1999.
- [69] M. Bernasconi, P.L. Silvestrelli, and M. Parrinello. *Phys. Rev. Lett.*, 81:1235, 1998.
- [70] R. Resta. *Phys. Rev. Lett.*, 80:1800, 1998.
- [71] N. Marzari and D. Vanderbilt. *Phys. Rev. B*, 56:12847–12865, 1997.
- [72] J. Cioslowski. *J. Am. Chem. Soc.*, 111:8333, 1989.
- [73] X. Gonze. *Phys. Rev. B*, 55:10337, 1997.
- [74] F. De Proft, J. M. L. Martin, and P. Geerlings. *Chem. Phys. Lett.*, 250:393, 1996.
- [75] A. Pasquarello and R. Car. *Phys. Rev. Lett.*, 79:1766, 1997.
- [76] A. Pasquarello and R. Resta. *Phys. Rev. B*, 68:174302, 2003.
- [77] J. Hutter et al. *CPMD, version 3.7.1*. IBM Research Division, IBM Corp and Max Planck Institute Stuttgart.
- [78] P. L. Silvestrelli and M. Parrinello. *Phys. Rev. Lett.*, 82:3308, 1999.
- [79] P. Tangney. *J. Chem. Phys.*, 124:044111, 2006.
- [80] P. Tangney and S. Scandolo. *J. Chem. Phys.*, 116:14, 2002.
- [81] J. M. Bowman, X. Zhang, and A. Brown. *J. Chem. Phys.*, 119:646, 2003.
- [82] M. Kaledin, A. Brown, A. Kaledin, and J. M. Bowman. *J. Chem. Phys.*, 121:5646, 2004.
- [83] J. Kohanoff. *Comp. Mat. Sc.*, 2:221, 1994.
- [84] G. Onida, W. Andreoni, J. Kohanoff, and M. Parrinello. *Chem. Phys. Lett.*, 219:1, 1994.
- [85] Pnina Dauber-Osguthorpe, David J. Osguthorpe, Peter S. Stern, and John Moulton. *J. of Comput. Phys.*, 151:169, 1999.
- [86] Richard B. Sessions, David J. Osguthorpe, and Pnina Dauber-Osguthorpe. *J. Phys. Chem.*, 99:9034, 1995.
- [87] Pnina Dauber-Osguthorpe and David J. Osguthorpe. *J. Comput. Chem.*, 14:1259, 1993.
- [88] Pnina Dauber-Osguthorpe and David J. Osguthorpe. *J. Am. Chem. Soc.*, 112:7921, 1990.
- [89] R. M. Stratt. *Acc. Chem. Res.*, 28:201, 1995.
- [90] M. Buchner, B. Ladanyi, and R. M. Stratt. *J. Chem. Phys.*, 97:8522, 1992.
- [91] M. Cho, G. R. Fleming, S. Saito, I. Ohmine, and R. M. Stratt. *J. Chem. Phys.*, 100:6672, 1994.
- [92] R. E. Larsen and R. M. Stratt. *J. Chem. Phys.*, 110:1036, 1999.
- [93] Ronald M. Levy, Olivia de la Luz Rojas, and Richard A. Friesner. *J. Phys. Chem.*, 88:4233, 1984.
- [94] R. A. Wheeler, H. Dong, and S. E. Boesch. *Chem. Phys. Chem.*, 4:382, 2003.
- [95] R. A. Wheeler and H. Dong. *Chem. Phys. Chem.*, 4:1227, 2003.
- [96] B. R. Brooks, D. Janežic, and M. Karplus. *J. Comp. Chem.*, 16:1522, 1995.
- [97] A. Amadei, A. B. M. Linssen, and H. J. C. Berendsen. *Proteins: Struct. Funct. Genet.*, 17:412, 1993.
- [98] F. Tama, F.X. Gadea, O. Marques, and Y.H. Sanejouand. *Proteins: Struc. Func. Genet.*, 41:1, 2000.
- [99] F. Tama and Y.H. Sanejouand. *Protein Eng.*, 14:1, 2001.
- [100] M. Schmitz and P. Tavan. *J. Chem. Phys.*, 121(24):12247–12258, 2004.
- [101] M. Schmitz and P. Tavan. *J. Chem. Phys.*, 121(24):12233–12246, 2004.
- [102] A. Strachan. *J. Chem. Phys.*, 120:1, 2004.
- [103] R. Kubo, M. Toda, and N. Hashitsume. *Statistical Physics II-nonequilibrium statistical mechanics*. Springer, Heidelberg, 2 edition, 1985.
- [104] J.C. Decius. *J. Chem. Phys.*, 17:1315, 1949.
- [105] P. Pulay and F. Fogarasi. *J. Chem. Phys.*, 96:2856, 1992.
- [106] Y. Morino and K. Kuchitsu. *J. Chem. Phys.*, 20:1809, 1952.
- [107] J.C. Whitmer. *J. Mol. Spec.*, 68(2):326, November 1977.
- [108] W.J. McCarthy, L. Lapinski, M. J. Nowak, and L. Adamowicz. *J. Chem. Phys.*, 108:10116, 1998.
- [109] B.L. Crawford and W.H. Fletcher. *J. Chem. Phys.*, 19:141, 1951.

Mode	Freq. (cm^{-1})	Intens. (km/mol)	Contributions from PED
gas phase cis-NMA			
Amide III	1255	58.9	C1 sym.bend.:47%, st.C2-N:22%, b.N-H:7%
C1 sym.bend	1322	217.5	C1 sym.bend.:98%
Amide II	1365	0.9	b.N-H:42 %,st.C=O:32%
Méthyl def.	1396	22.3	C1 asym.bend.:61%, C3 sym.bend.:28%
Méthyl def.2	1413	20.7	C1 asym.bend.:46%, C3 asym.bend.:39%
Amide I	1598	451.4	st C=O: 92%, b. N-H: 4%
gas phase trans-NMA			
Amide III	1179	63.7	st.C2-N:32%, b.N-H:30%, st.C=O:12%
C1 sym.bend.	1320	33.0	C1 sym.bend.:95%, C1 asym.bend.:3%
C3 sym.bend.	1363	36.7	C3 sym.bend.:77%, b.N-H:11%
Méthyl bend.	1394	40.2	C1 asym.bend.:85%, b.N-H:4%
Amide II	1420	37.6	b.N-H:42%, C3 sym.bend.:17%, st.C2-N:10%
Amide I	1596	330.7	st.C=O:95%, b.N-H:2%
solvated trans-NMA			
Amide III	1243	124.8	st.C2-N:38%, b.N-H:35%, st.C=O:12%,
C1 sym. def.	1349	47.3	C1 sym.bend.:93%
C3 sym. def.	1374	99.2	C3 sym.bend.:79%, b.N-H:6%
Amide II	1493	478.9	b.N-H:48%, st.C2-N:24%, C3 sym.bend.:13%
Amide I	1515	235.9	st.C=O:54%, b.N-H:38%

Table 1. PED and IR intensities of localized modes between 1100 and 1700cm^{-1} for gas phase Cis- and Trans-NMA and solvated Trans-NMA, extracted from first-principle Molecular Dynamics simulations at 20 K (gas phase) and 300 K (liquid phase). Effective modes are defined using Pulay's non redundant internal coordinates [19]. st: stretch, b: bend, sym bend: symmetric bend, asym bend: asymmetric bend.

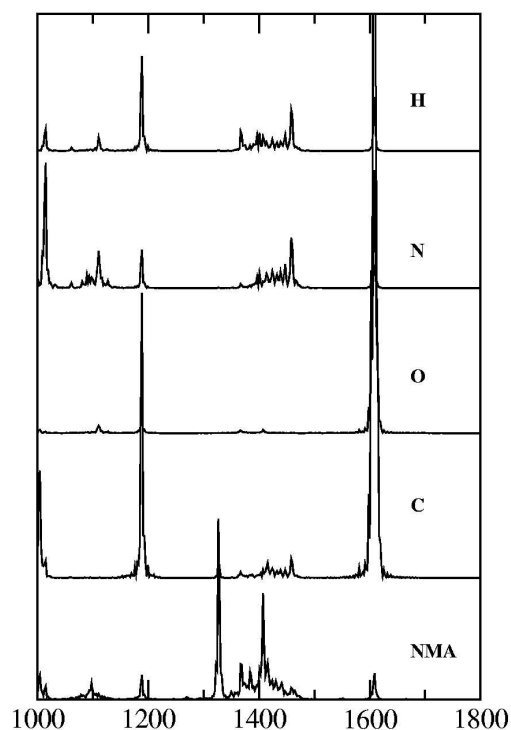


Figure 1. Vibrational density of states (VDOS) of isolated Trans-NMA (N-Methyl-Acetamide) obtained in a Car-Parrinello molecular dynamics at 20 K. The total VDOS is displayed at the bottom of the figures and the various contributions from individual atoms of the peptidic group (H, N, C, O) are given above. C and O respectively stand for the carbonyl carbon and oxygen atom of NMA, N and H for the amide nitrogen and hydrogen atoms.

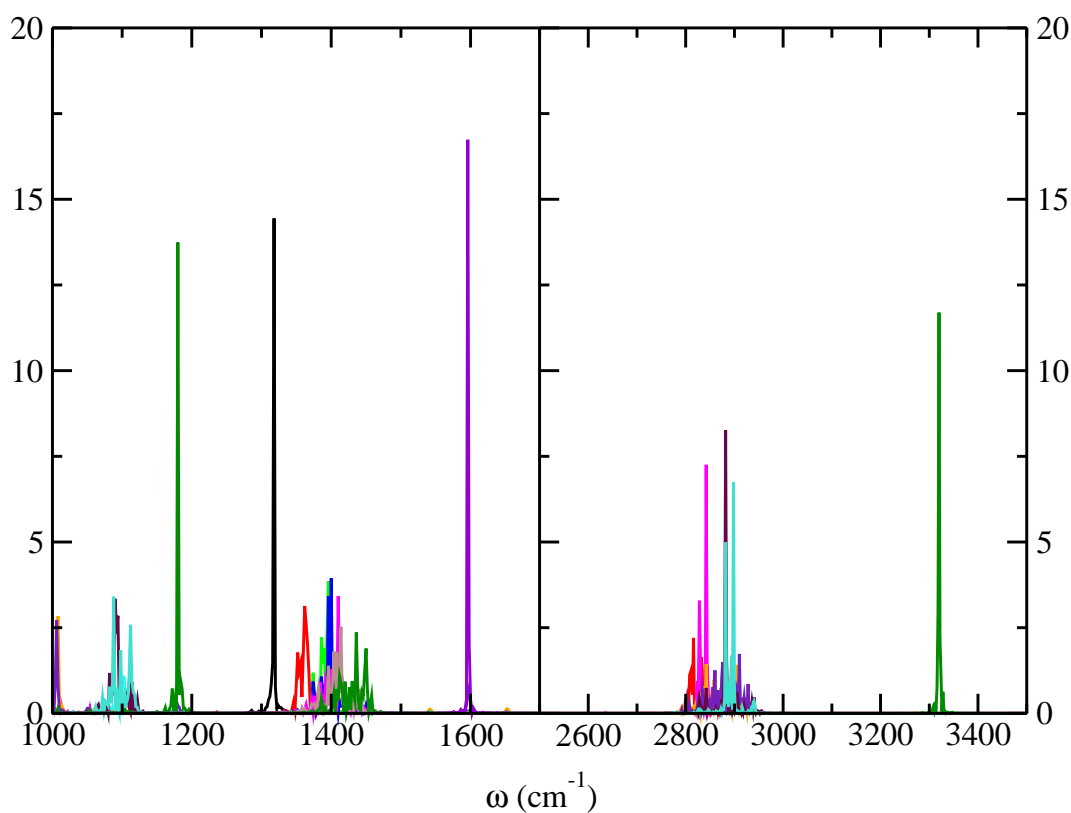


Figure 2. Power spectra of the localized modes for gas phase Trans-N-Methyl-Acetamide at 20 K.

The effective normal modes are computed with the formalism described in section 5 using non-redundant internal coordinates as defined by Pulay [19].

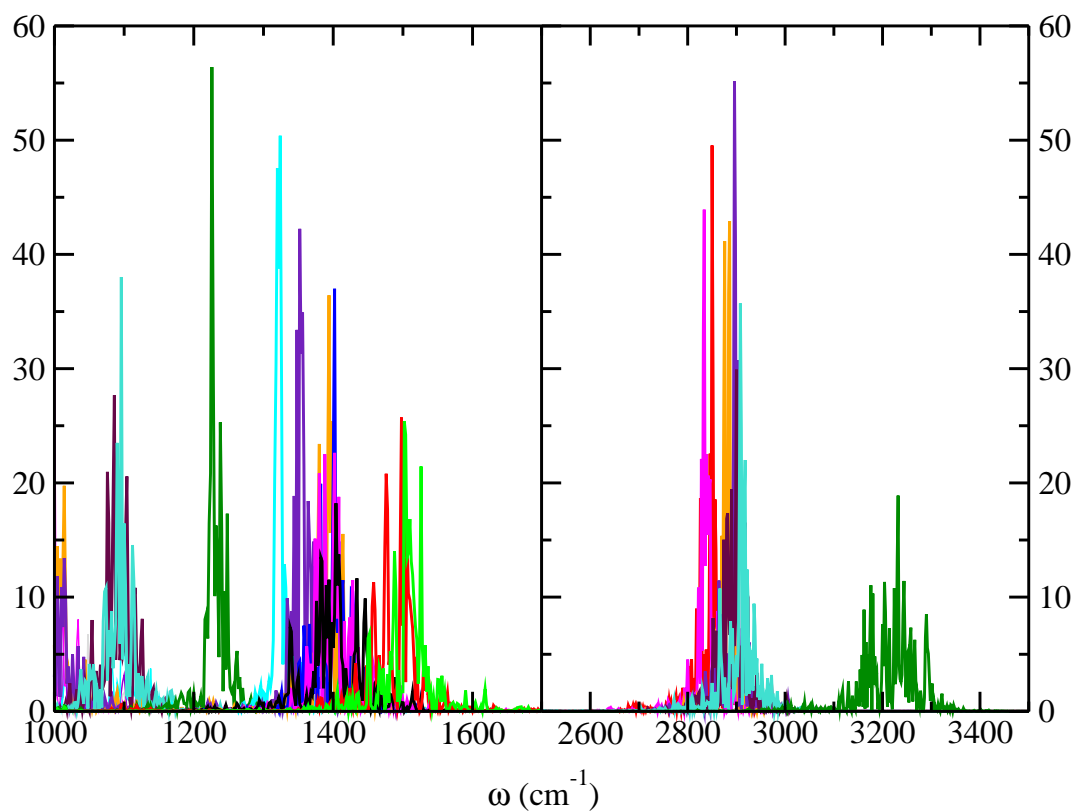


Figure 3. Power spectra of the localized modes for solvated Trans-N-Methyl-Acetamide at 300 K. The effective normal modes are computed with the formalism described in section 5 using non-redundant internal coordinates as defined by Pulay [19].

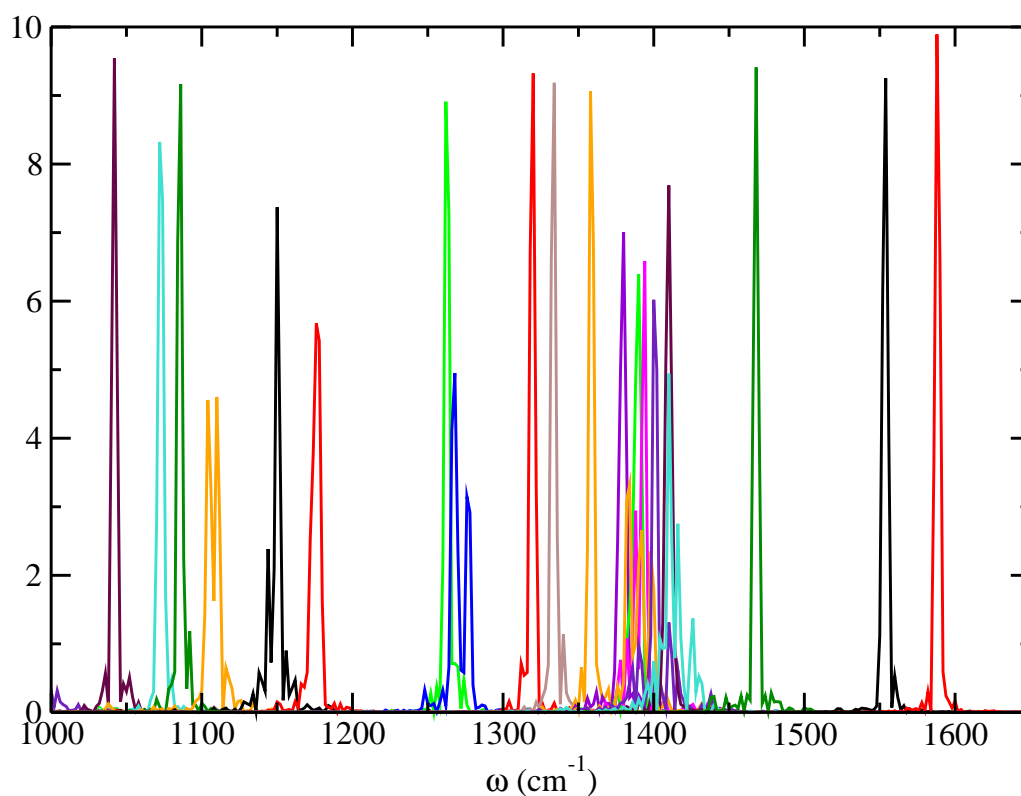


Figure 4. Power spectra of the localized modes for gas phase alanine di-peptide in the C7eq conformation, from the gas phase Car-Parrinello dynamics at 20 K. The effective normal modes are computed with the formalism described in section 5 using non-redundant internal coordinates as defined by Pulay [19].

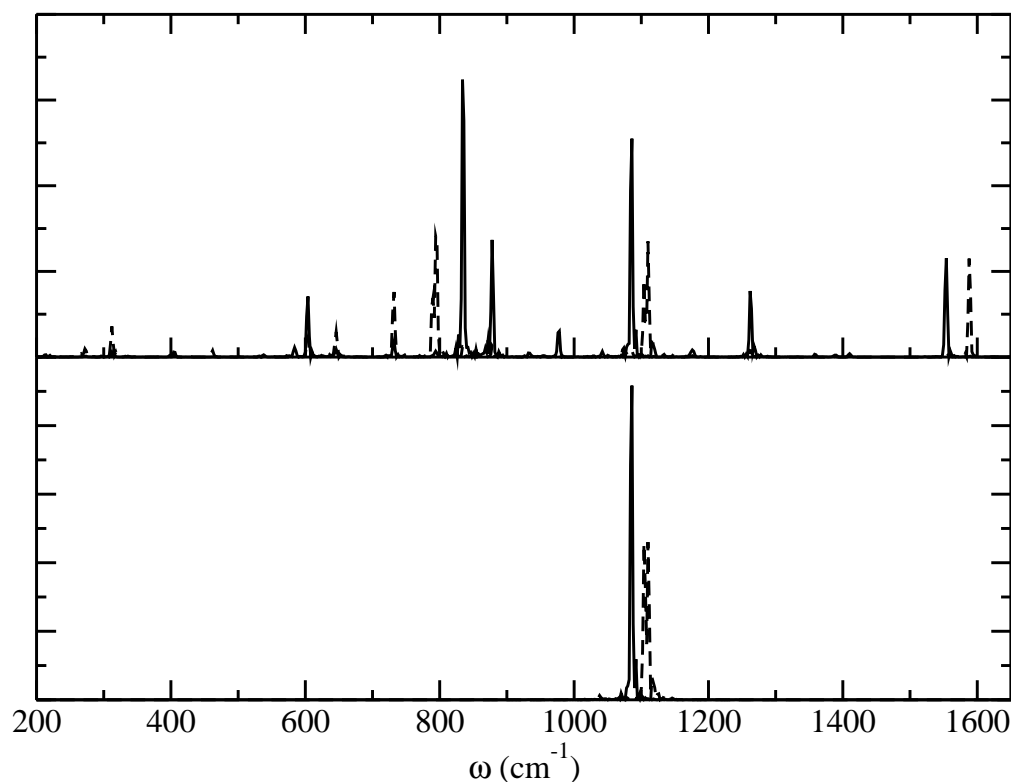


Figure 5. Top: VDOS of the two amide N-H groups of an alanine dipeptide calculated with the usual definition in equation (15) restricted to the N and H atoms only of each N-H group, see section 4. One N-H group is identified with the black line, the other with the dashed black line. See text for the related discussion. Bottom: Power spectra of the effective normal modes associated to the N-H Amide III modes after the localization transformation presented in section 5 using non-redundant internal coordinates as defined by Pulay [19]. See text for the related discussion.

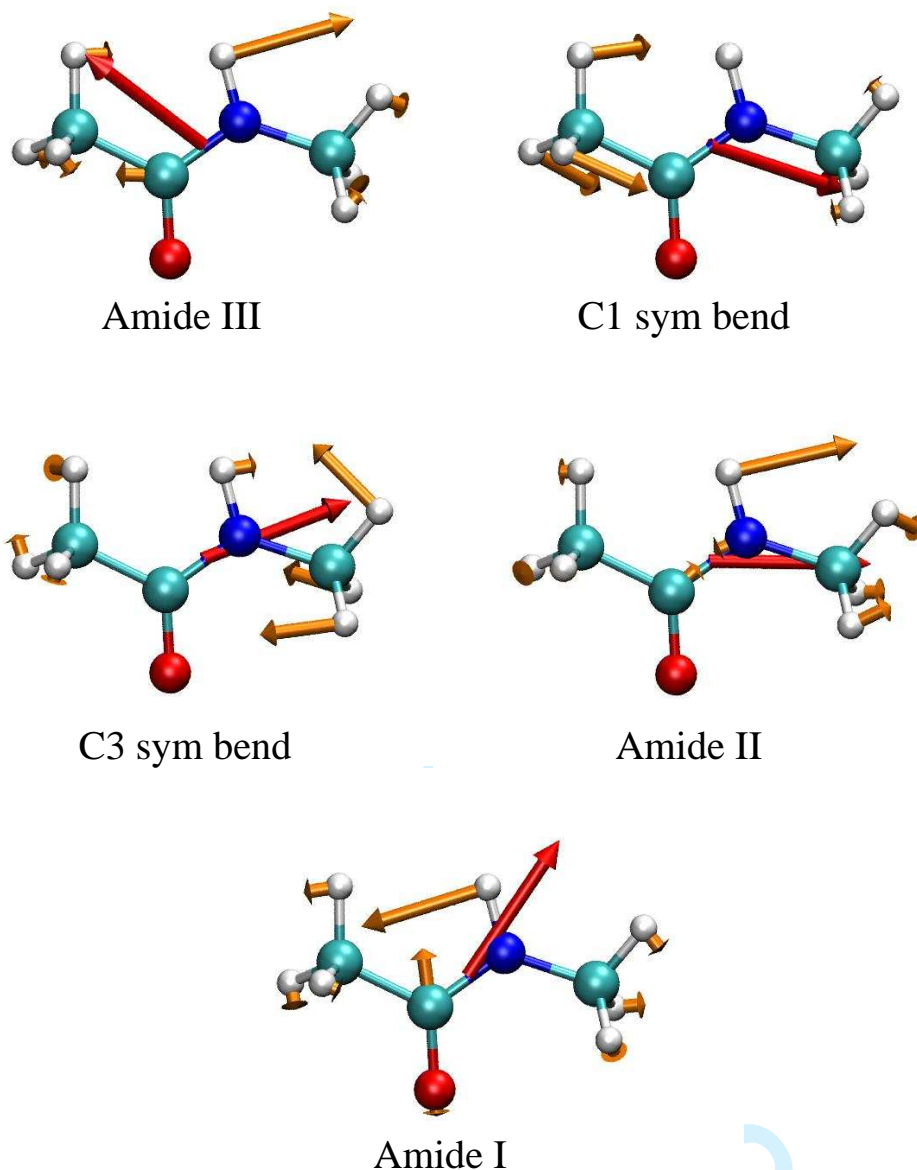


Figure 6. Atomic displacements (orange arrows) for some selected effective normal modes of solvated Trans-NMA. For each mode, the transition dipole as defined by eq. 39 is also shown as a red arrow.

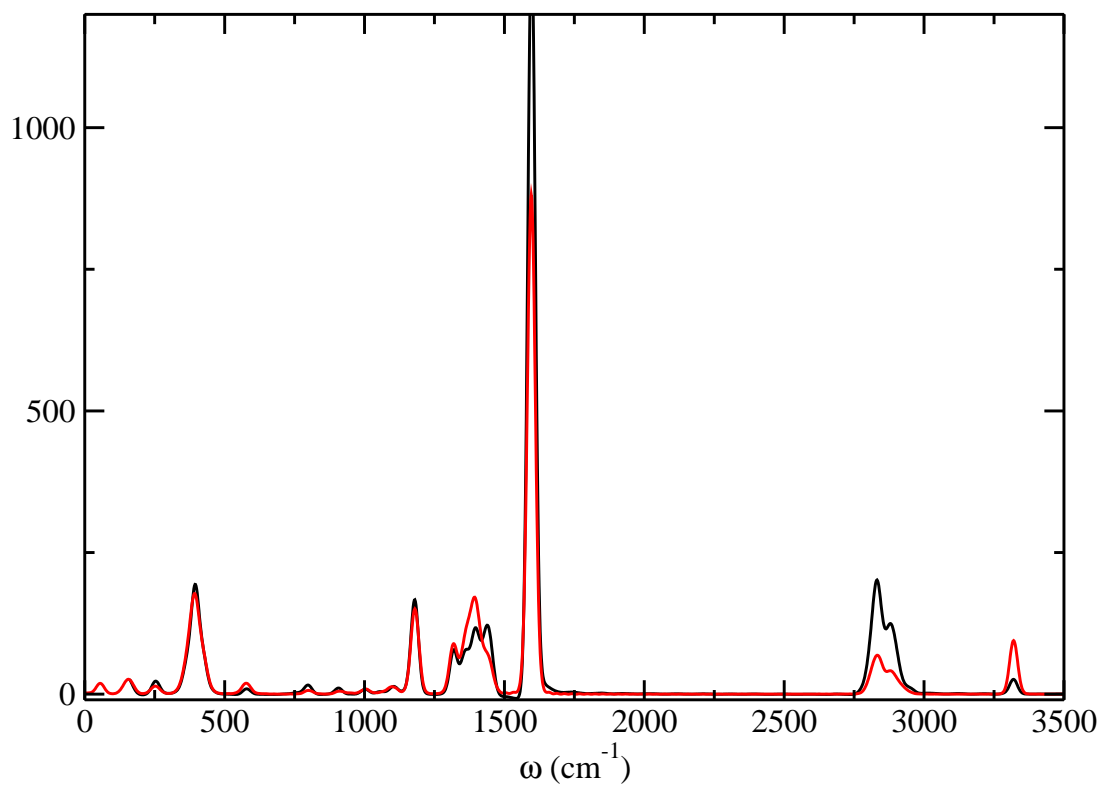


Figure 7. Comparison between the infrared spectrum calculated through the Fourier transform of the dipole correlation function (equation (8), black line) and the Fourier transform of the localized normal modes (equation (44), red line) making use of APT tensors (APT tensor calculated at the first configuration of the dynamics). Comparison presented for the Car-Parrinello molecular dynamics of Trans-NMA in the gas phase at 20 K.

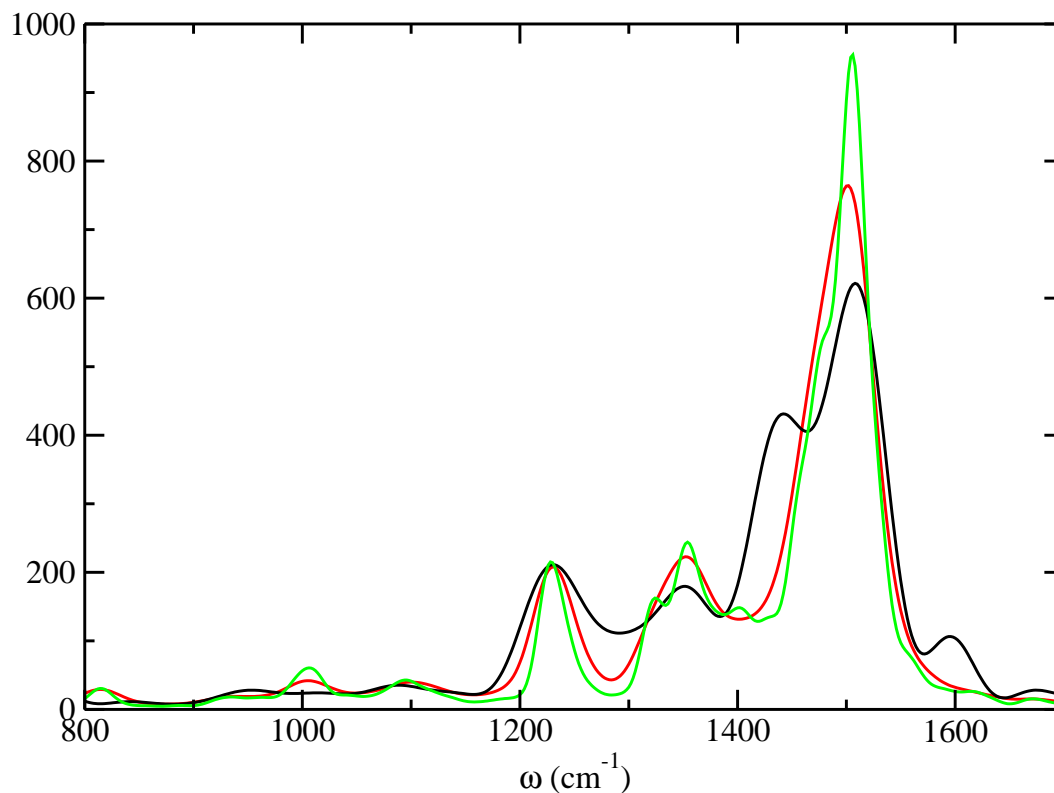


Figure 8. Comparison between the infrared spectrum calculated through the Fourier transform of the dipole correlation function (equation (8), black line) and the Fourier transform of the localized normal modes (equation (44), red and green lines) making use of APT tensors (red: APT tensor sampled over one pico-second trajectory, green: APT tensor calculated at the first configuration of the dynamics). Comparison presented for the Car-Parrinello molecular dynamics of Trans-NMA immersed in liquid water at 300 K.

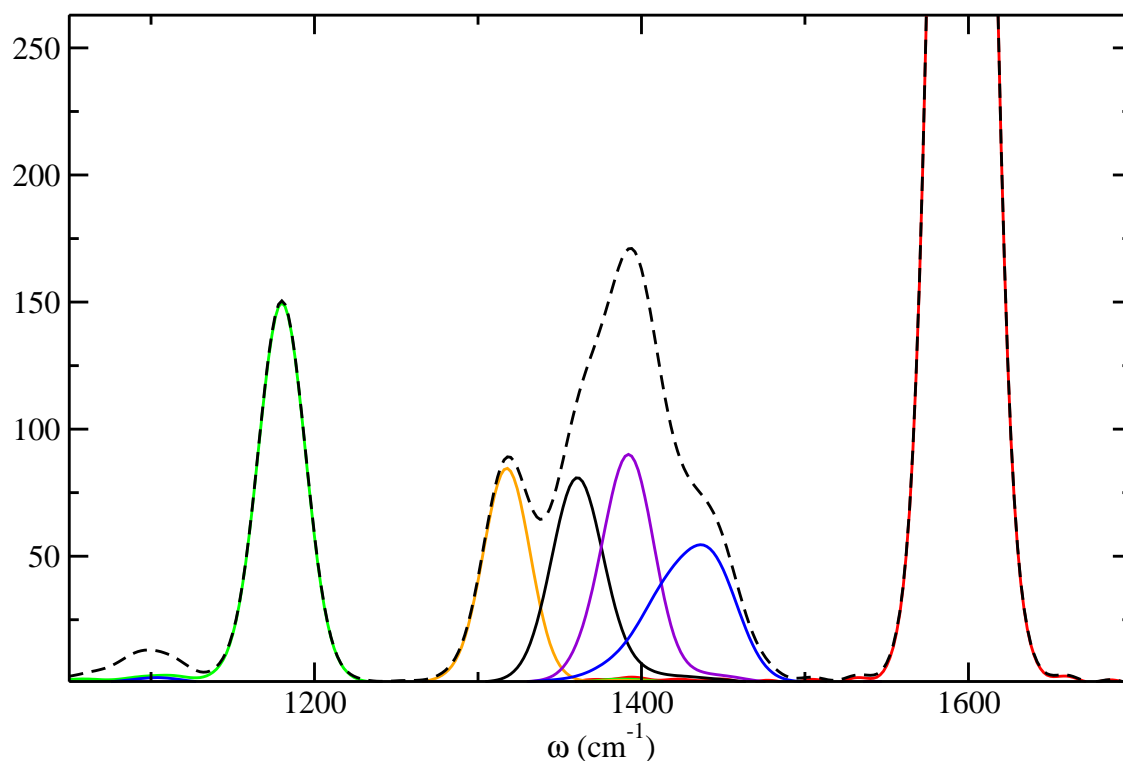


Figure 9. Decomposition of the infrared spectrum intensity of Trans-NMA from the 20 K simulation in the gas phase. Decomposition into the different contributions arising from the Amide normal modes in the 1000-1800 cm^{-1} frequency domain: Dashed Black: total IR spectrum; Red: Amide I mode; Blue: Amide II mode; Orange, black and purple: methyl groups C-H bending; green: Amide III mode.

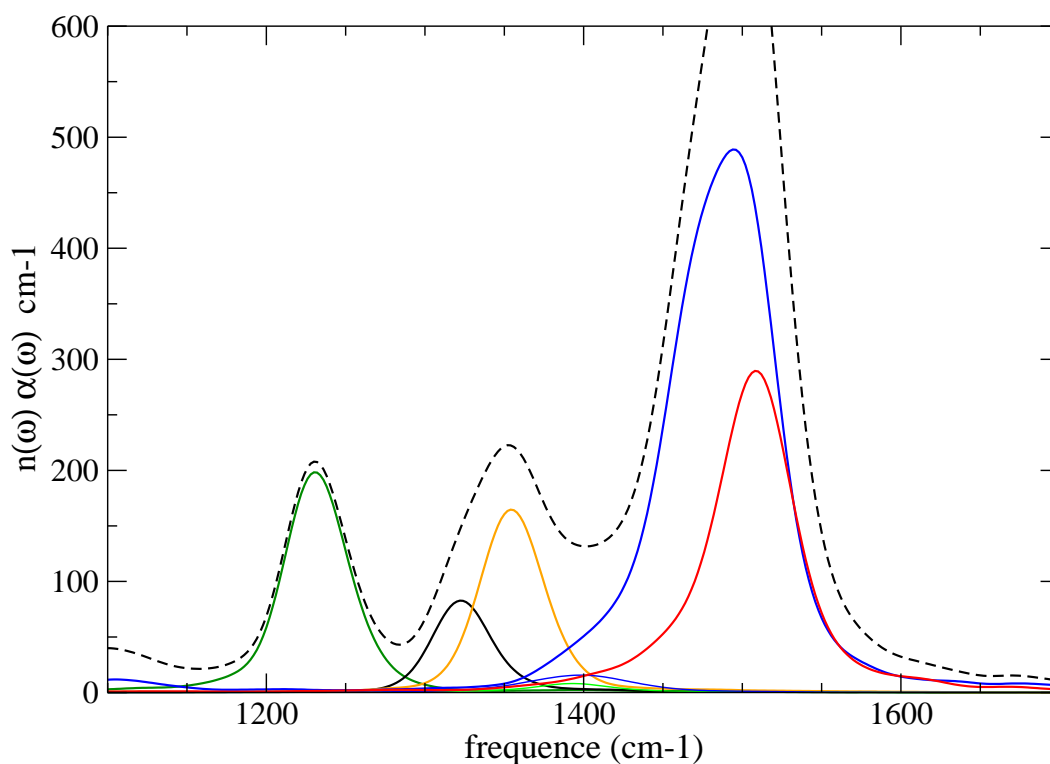


Figure 10. Decomposition of the infrared spectrum intensity of Trans-NMA solute from the 300 K aqueous Trans-NMA simulation. Decomposition into the different contributions arising from the Amide normal modes in the 1000-1800 cm^{-1} frequency domain: Dashed Black: total IR spectrum; Red: Amide I mode; Blue: Amide II mode; Orange and black: methyl groups C-H bending modes, respectively from the C-Terminal side and N-Terminal side of the molecule; green: Amide III mode.

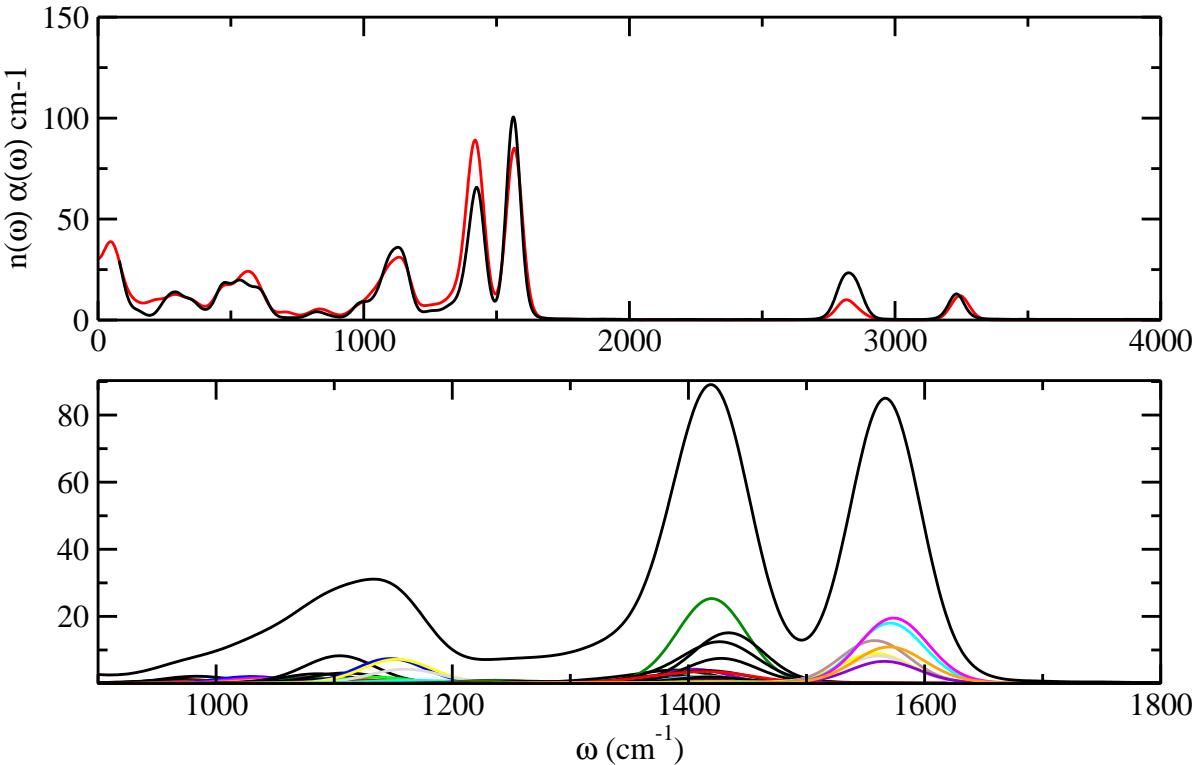


Figure 11. Gas phase octa-alanine at 20 K. Top: Comparison between the infrared spectrum calculated through the Fourier transform of the dipole correlation function (equation (8), black line) and the Fourier transform of the localized normal modes (equation (44), red line) making use of APT tensors (APT tensor calculated at the first configuration of the dynamics). Bottom: Decomposition of the infrared spectrum intensity into the different contributions arising from the Amide normal modes in the 1000-1800 cm^{-1} frequency domain: Dashed Black: total IR spectrum; Color curves: individual effective normal modes.

Infrared spectroscopy in the gas and liquid phase from first principle molecular dynamics simulations - Application to small peptides

Marie-Pierre Gaigeot^{1,2,*}, Michaël Martinez^{2,†} and Rodolphe Vuilleumier²

¹ LAMBE Laboratoire Analyse et Modélisation pour la Biologie et l'Environnement, UMR-CNRS 8587

Université d'Evry val d'Essonne, Blvd F. Mitterrand, Bât. Maupertuis, F-91025 Evry, France, gaigeot@ccr.jussieu.fr

² LPTMC Laboratoire de Physique Théorique de la Matière Condensée, UMR-CNRS 7600

Tour 24-25, 2^{ème} étage, c.c. 121

Université Pierre et Marie Curie, 4 place Jussieu, F-75005 Paris, France, vuilleum@lptmc.jussieu.fr

(Received 00 Month 200x; in final form 00 Month 200x)

We discuss the applicability of finite temperature Car-Parrinello molecular dynamics simulations for the calculation of infrared spectra of complex molecular systems, either in the gas phase or in the condensed phase, taking examples from the infrared spectroscopy of N-methylacetamide and small peptides. Band assignments for the simulation is still challenging and we introduce here a general method for obtaining effective normal modes of molecular systems from Molecular Dynamics simulations. The effective normal modes are defined as linear combination of internal coordinates such that the power spectra of these modes are as localized as possible in frequency. We further define band intensities for these modes from different levels of approximation of the infrared spectrum. Applications of this approach for assigning infrared bands from first-principle molecular dynamics simulations are presented for N-methylacetamide in gas phase and in solution, for the gas phase alanine dipeptide and the gas phase octa-alanine peptide.

1 Introduction

Vibrational spectroscopy (Infrared and Raman) has been recognized for long as a fundamental tool for the characterization and understanding of the chemical state of a molecule, the structure and dynamics of the molecule and its environment. Vibrational spectroscopy is in particular commonly applied for the characterization of secondary structures of peptides and proteins in solution or membranes [1–4]. As an illustration, structural motifs such as helices or beta sheets can be clearly identified from each others by changes in frequency positions, shapes and intensities of the infrared or Raman bands [2, 5, 6].

Linear absorption Infrared spectroscopy has been successfully applied for the last two decades on all kinds of molecular systems in the liquid, solid and gas (matrix) phases. But the past decade has seen the developement of new vibrational spectroscopy methods. Pump-probe non linear 2-dimensional infrared spectroscopy 2D-IR [7–12] is one of these methods. It has been developed in order to circumvent the limitations of linear IR, in particular bringing forth the couplings between modes and the extra-information it gives in the understanding of molecular structures. The advance in 2D-IR is the time dependence recording of the infrared fingerprints: dynamics of a molecular system is thus directly at hand with these experiments, in real time. Up to now, the methodology has been principally developed on the Amide I mode (stretching C=O cabonyl) and mainly applied on biomimetic organic molecules such as peptide models in the liquid phase. Same developments are being made with multi-dimensional Raman spectroscopy [13, 14]. Gas phase spectroscopy has also been revolutionarized during the past decade, going far beyond the low-temperature matrix gas phase spectroscopy. Hence, gas phase mass spectrometry and infrared vibrational spectroscopy have been gathered in a set-up which thus couples Mass Spectrometry (MS) with Infrared Multi Photon Dissociation (IR-MPD), as devised at FELIX [15] or CLIO [16] in Europe. This

*Corresponding authors: gaigeot@ccr.jussieu.fr & vuilleum@lptmc.jussieu.fr

†Present adress: Molecular & Cellular Modeling Group, EML Research gGmbH, 69118 Heidelberg, Germany

is an action spectroscopy, where the fragment ion yields are recorded as a function of the IR excitation energy. Conditions necessary for IR-MPD of gas phase molecules are fulfilled by using an infrared free electron laser (FEL). Thermalization of the produced ions is an important aspect of these experiments. Though the ions are initially produced at high internal energies, their trapping in an ICR (Ion Cyclotron Resonance) cell during a few seconds before exposure to the IR radiation ensures their relaxation and thermalization at room temperature. This device therefore gives the IR spectroscopy of trapped gas phase ions at room temperature, close enough to relevant physiological temperatures of biomolecules. See for instance refs. [17, 18] for more information on the subject.

We are thus able to vibrationally probe molecules either in the gas phase or in the liquid phase, at finite temperature: a direct comparison between the gas phase and liquid phase vibrational patterns can therefore be conducted, *at the same finite temperature*. Environmental effects, such as the aqueous hydrogen bonding to biomolecules, can be directly probed and characterized from this comparison.

Assignment of vibrational modes is an important step in the structural and dynamical interpretation of the spectral fingerprints. Here is where modelling and computation can be most helpful, in particular when a full spectrum, including positions, shapes and intensities, can be computed and directly compared to experimental data. This however necessitates accurate intra- and inter-molecular potentials, able to both describe the typical intramolecular motions of the molecule under study but also the induced shifts due to conformational changes or due to the interaction with the environment. Such potential should then correctly incorporate anharmonic terms and polarisation effects, in particular at finite temperature.

This is in a way exactly the same statement as in the seminal work of Péter Pulay [19] ! Since then, great progress has though been made on the computation and interpretation of vibrational spectra of molecules of reasonable size and complexity, such as bio-organic molecules which are of main interest for many research groups nowadays.

The conventional way of calculating infrared spectra principally involves performing ab initio chemistry calculations [18, 20]. This is the traditional approach for dealing with small organic molecules. In such a scheme, geometry optimizations at 0 K are performed and vibrational analyses in the harmonic approximation are subsequently done for each equilibrium conformation. A match between experimental and calculated IR patterns obtained for the different conformers is generally checked in terms of positions and relative intensities of IR bands. The bottleneck of such theoretical calculations is mainly the determination of all the representative equilibrium conformations of the molecule. This is a tremendous amount of work for flexible and floppy molecules. A small di-peptide already possesses several conformers (representative of those encountered in larger peptides) which are generally lying within a small energy range. At room temperature, one can thus expect conformational isomerisation dynamics to take place. In other words, the peptide does possess enough internal energy to explore different zones of its potential energy surface (PES), particularly if finite temperature IR-MPD gas phase experiments are involved. Therefore, different isomeric conformations of the molecule can be accessible at that temperature and will most likely play a role on the peptide IR spectroscopy. Another limitation (and not the least) of quantum chemistry calculations is the modeling of the environment surrounding the vibrationally probed solute molecule, water being the most ubiquitous solvent. Properly characterising the surrounding hydration shells is indeed out of reach of such calculations, where only a few water molecules that belong to the first hydration shell and are hydrogen bonded to the main hydrophilic sites of the solute can be incorporated into the calculation.

Taking into account the dynamics of the molecule and its consequences on the measured properties can only be achieved through molecular dynamics simulations (MD). MD simulations are also the only approach capable of modeling condensed phase systems, such as homogeneous liquids or a solute immersed in liquid water (or any other solvent environment). Lately, Density Functional Theory (DFT) based Molecular Dynamics has proven successful for the computation of vibrational spectra at finite temperature. This development is the result of the latest advances in density functional theory (DFT) based molecular dynamics simulation ("Car-Parrinello" [21]), brought about by the implementation of the modern theory of polarization and the maximally localized Wannier functions. In Car-Parrinello molecular dynamics simulations the forces acting on the atoms are obtained at each time-step of the simulation by an electronic structure calculation in the DFT framework. This thus combines accurate description of forces in many

environments and situations, while providing a description of the structure and dynamics of a system in its fluctuating environment at finite temperature. Since the pioneering work of Silvestrelli *et al.* for the computation of the infrared spectrum of liquid water [22], DFT-based MD has been applied successfully to the calculation of IR spectra of pure liquids [23,24] or of small biological molecules in aqueous solution [25–29], molecular crystals [30] or Raman spectra in the condensed phase [31–33]. As we have demonstrated in recent studies of the IR absorption of uracil [34], N-Methyl-Acetamide [35], and the protonated alanine peptide [18,36], either in liquid water or in the gas phase, it is now possible to perform a detailed infrared calculation for models of bio-organic molecules in finite temperature bulk solution with all water molecules treated at the same level of theory as the solute, as well as in finite temperature gas phase.

An accurate calculation of infrared spectra is one issue to achieve, the assignment of the active bands into individual atomic displacements or vibrational modes is another one, and this issue is essential to the understanding of the underneath molecular structural and dynamical properties.

We have recently developed a systematic method for decomposing vibrational spectra as an approximate sum of effective normal mode contributions [37]. We have defined these effective normal modes from the data of the Fourier Transform of Velocity Correlation Functions so that they can be extracted directly from the MD trajectory. In this framework, the effective normal modes are linear combinations of atomic displacements constructed such that the corresponding power spectra is as localized as possible in frequency. There is no loss of information from this transformation which is just a change of coordinates.

The present paper aims at demonstrating the applicability of *ab initio* molecular dynamics simulations (such as the Car-Parrinello approach in the DFT framework) for the calculation of infrared spectra of bio-organic molecules (in relation to gas phase and liquid phase finite temperature experiments) and our methodology for the extraction and assignment of vibrational modes in terms of internal coordinates. We will concentrate on the infrared spectroscopy of peptide models, either in the gas phase or in the liquid phase, in relation with finite temperature linear IR or IR-MPD experiments. We will begin with a summary review of static *versus* dynamics calculations of infrared spectra, presenting the advantages and limitations of each method. The localization method we have developed will be described next, as well as a way for the calculation of infrared intensities in our framework. Applications will be presented, mainly on the N-Methyl-Acetamide molecule, Alanine di-peptides and the octa-alanine peptide chain, in the gas phase and in the liquid phase. Ongoing further developments will be mentioned at the end of the paper.

2 Static *versus* dynamic *ab initio* calculations of infrared spectra

The basic definition of an infrared spectrum calculation lies in the Fermi Golden Rule which states that the infrared spectrum is calculated through the following expression [38]:

$$I(\omega) = 3 \sum_i \sum_f \rho_i |\langle f | \mathbf{E} \cdot \mathbf{M} | i \rangle|^2 \delta(\omega_{fi} - \omega) \quad (1)$$

where \mathbf{E} is the applied external field vector, \mathbf{M} is the dipole vector of the molecular system, $|i\rangle$ and $|f\rangle$ are respectively the initial and final vibrational states of the system. These are eigenstates of the system excluding the radiation. ρ_i is the density of molecules in the initial vibrational state $|i\rangle$. ω is the reciprocal wave-number (in cm^{-1}) and ω_{fi} is the reciprocal wave-number associated with the transition between the initial and final vibrational states of the system.

• Infrared spectroscopy with static *ab initio* calculations

The root of vibrational spectra calculations is solving the Schrödinger vibrational equation [39,40]. It is expressed within the harmonic normal mode coordinates q_k ($k = 1 \cdots 3N - 6$, where N is the number

of atoms of the investigated molecular system) in order to be analytically and numerically solvable very easily. With these coordinates, the kinetic energy T and potential energy V that appear in the Schrödinger vibrational equation are simply expressed as:

$$T = \frac{1}{2} \sum_{k=1}^{3N-6} \dot{q}_k^2 \quad V = \frac{1}{2} \sum_{k=1}^{3N-6} \lambda_k q_k^2$$

where λ_k is the frequency associated to the normal mode q_k , and the Schrödinger vibrational equation is:

$$-\frac{\hbar^2}{8\pi^2} \sum_{k=1}^{3N-6} \frac{\partial^2 \Psi_V}{\partial q_k^2} + \frac{1}{2} \sum_{k=1}^{3N-6} \lambda_k q_k^2 \Psi_V = E_V \Psi_V \quad (2)$$

where Ψ_V is the vibrational wavefunction and E_V the corresponding energy; \hbar is the Planck's constant. Due to the normal modes decouplings, the vibrational energy E_V and vibrational wavefunction Ψ_V can be decomposed as

$$E_V = E(q_1) + E(q_2) + \dots + E(q_{3N-6}) \quad \Psi_V = \Psi(q_1)\Psi(q_2) \dots \Psi(q_{3N-6})$$

so that equation (2) is satisfied if the following $3N-6$ decoupled equations are satisfied

$$-\frac{\hbar^2}{8\pi^2} \frac{\partial^2 \Psi(q_k)}{\partial q_k^2} + \frac{1}{2} \lambda_k q_k^2 \Psi(q_k) = E_k \Psi(q_k) \quad \forall k = 1 \dots 3N-6 \quad (3)$$

Solution of the Schrödinger vibrational equation of a harmonic oscillator q_k (equation 3) is well-known, with the vibrational energy $E_k = (v_k + 1/2)\hbar\nu_k$ (where v_k is a quantum number associated with the normal frequency ν_k), and the vibrational wavefunction $\Psi_{v_k}(q_k) = C_k \exp(-1/2\gamma_k q_k^2) H_{v_k}(\gamma_k q_k)$ where C_k is a normalising factor, $\gamma_k = 4\pi^2\nu_k/\hbar$ and $H_{v_k}(\gamma_k q_k)$ is the Hermite polynomial of degree v_k in q_k . The harmonic oscillator wavefunctions are orthonormalised, i.e.:

$$\int dq_k \Psi_{v_k}^*(q_k) \Psi_{v'_k}(q_k) = \delta_{v_k v'_k} \quad (4)$$

The intensity of a spectral line is the probability of the transition which gives rise to the line: the coefficient of absorption of light of frequency $\nu_{v_k v'_k}$ for infrared spectroscopy is given by the Einstein coefficient [41]

$$B_{v_k v'_k} = \frac{8\pi^3}{3\hbar^2} \left[|\mu_{X v_k v'_k}|^2 + |\mu_{Y v_k v'_k}|^2 + |\mu_{Z v_k v'_k}|^2 \right]$$

where μ_X , μ_Y , μ_Z are the components of the transition dipole moment:

$$\mu_{\alpha V V'} = \int d\tau \Psi_V^* \mu_{\alpha} \Psi_V \quad \alpha = X, Y, Z \quad (5)$$

with μ_{α} the α^{th} (X, Y, Z) component of the electric dipole moment of the molecular system. Ψ_V and $\Psi_{V'}$ are the total vibrational wavefunctions for states V and V' . The integral is taken over the configuration space of the system.

By application of the electric harmonic approximation (second harmonic approximation), each μ_α component can be written:

$$\mu_\alpha = \mu_\alpha^{(0)} + \sum_{k=1}^{3N-6} \mu_\alpha^{(k)} q_k$$

where $\mu_\alpha^{(0)}$ is the permanent dipole moment of the molecule evaluated at the equilibrium geometry, and $\mu_\alpha^{(k)} = \left(\frac{\partial \mu_\alpha}{\partial q_k} \right)_0$ is also evaluated at the equilibrium geometry. Equation (5) now reads:

$$\mu_{\alpha V V'} = \mu_\alpha^{(0)} \int d\tau \Psi_{V'}^* \Psi_V + \sum_{k=1}^{3N-6} \left(\frac{\partial \mu_\alpha}{\partial q_k} \right)_0 \int d\tau \Psi_{V'}^* q_k \Psi_V \quad (6)$$

which gives rise to the well-known infrared selection rules of harmonic oscillators (with orthonormalized wave-functions, eq. 4), i.e. $\mu_{\alpha V V'} \neq 0$ if $v_{l'} = v_l$ for all vibrational levels except $v_{k'}$ and v_k for which $\int dq_k \Psi_{v_{k'}}^* q_k \Psi_{v_k}$ is different from zero if $v_{k'} = v_k \pm 1$ (absorption or emission of one photon from two consecutive vibrational levels). The permanent dipole $\mu_\alpha^{(0)}$ does not play any role in the intensity of the spectral lines.

The absolute intensity of an active infrared transition is given by [39]

$$\kappa = \frac{8\pi^3}{3ch} \nu_{V V'} (N_{V'} - N_V) \left[|\mu_{X V V'}|^2 + |\mu_{Y V V'}|^2 + |\mu_{Z V V'}|^2 \right]$$

where N_V and $N_{V'}$ are the number of molecules per unit volume respectively in the state V and V' , and c is the speed of light. In the double harmonic approximation, κ is thus proportional to the sum of the transition dipole derivatives with respect to the harmonic normal modes, i.e.

$$\kappa = \sum_{\alpha=x,y,z} \sum_{k=1}^{3N-6} \left| \left(\frac{\partial \mu_\alpha}{\partial q_k} \right)_0 \right|^2 \delta_{v_{k'}, v_k \pm 1}. \quad (7)$$

In the double harmonic approximation, the Fermi Golden Rule thus reduces to the calculation of the normal modes of the molecular system in its equilibrium geometry and to the calculation of the transition dipole derivatives $\left(\frac{\partial \mu_\alpha}{\partial q_k} \right)_0$. Normal modes q_k give the frequency value at which a $v_k = 0 \rightarrow v_{k'} = 1$ transition takes place and the associated atomic movements, while dipole derivatives give the infrared intensities associated to each mode. The investigated infrared spectroscopy is related to the absorption of one photon from $v = 0$ to $v = 1$ harmonic vibrational transitions.

This is achieved with geometry optimization and harmonic frequency calculations performed with quantum chemistry calculations. The DFT (Density Functional Theory) framework has the advantage of allowing the study of molecular systems of tens to hundreds of atoms routinely on most of the computer platforms. The conventional way of proceeding is to perform a systematic search for the different possible molecular conformations of lowest energy on the potential energy surface, followed by harmonic frequency calculations of the different identified conformers and the related transition dipole derivatives. Frequencies are usually scaled [20]: the main purpose of the scaling is to compensate for the double harmonic approximations (potential energy surface and dipole moment expression) as well as for the level of the ab initio calculations. The calculated absorption spectra are finally convoluted with a Gaussian or Lorentzian band profile, adjusted in agreement with experimental conditions.

Obviously, the main drawbacks of this approach are *i*) the search for the minima of lowest energy on the potential energy surface, which can be tricky for floppy molecules, *ii*) the double harmonic approximations,

iii) as well as the inherent 0 K temperature associated to the search of the minima on the potential energy surface. This is in particular a very crude approximation for floppy molecules which can undergo conformational dynamics at finite temperature, as should occur in the infrared experiments we will be comparing our calculations to, which are typically performed at temperatures around 300 K. We will come back on these points later on.

• Infrared spectroscopy with ab initio molecular dynamics simulations

Within Linear Response Theory [38,42], the Fermi Golden Rule formula (equation 1) can be rewritten as the Fourier Transform of the dipole moment time correlation function:

$$I(\omega) = \frac{2\pi\beta\omega^2}{3cV} \int_{-\infty}^{\infty} dt \langle \mathbf{M}(t) \cdot \mathbf{M}(0) \rangle \exp(i\omega t) \quad (8)$$

where $\beta = 1/kT$, c is the speed of light in vacuum, V is the volume. The angular brackets indicate a statistical average of the correlation of the dipole moment \mathbf{M} of the absorbing molecular system. This average is taken in the absence of the applied external field. In this formula, we have taken into account a quantum correction factor (multiplying the classical line shape) of the form $\beta\hbar\omega/(1 - \exp(-\beta\hbar\omega))$, which was shown by us and others to give the most accurate results on calculated IR amplitudes [34,43,44]. For a complete discussion on quantum corrections, we refer the reader to refs. [45,46].

Equation (8) gives the whole infrared spectrum of a molecular system with *one single calculation*, i.e. the band positions, the band intensities and the band shapes, through the fourier transform of a time correlation function. There are no approximations made apart from the hypothesis of linear response theory, i.e. a small perturbation on the absorbing molecular system from the applied electric external field.

This is the standard way used in statistical mechanics for calculating the infrared spectrum of a molecular assembly of molecules (isolated molecule, liquids, solutes in the liquid phase, solids) [38,47–55]. Molecular dynamics simulations are adapted to the calculation of the evolution in time of the dipole moment of the system, and therefore its time correlation. We stress again that the only hypothesis on infrared spectra calculations through equation (8) corresponds to the linear response theory in which the applied external electric field is a small perturbation to the absorbing molecular system. In particular, in this kind of calculation there are no harmonic approximations made, be they on the potential energy surface or on the dipole moment, contrary to the static calculations previously described.

The main advantages of the molecular dynamics approach for the calculation of infrared spectra can be summarized as follows.

i) Dynamics simulations are performed at finite temperature; this will mainly be room temperature in our calculations as they will be compared with experiments typically done at this temperature. At finite temperature, there can be a conformational dynamics between different isomeric conformations of the absorbing molecules of interest. All conformations populated when going from one basin to the others on the potential energy surface are thus taken into account in the calculation of the infrared spectrum. This population dynamics therefore gives rise to a natural broadening of the calculated IR active bands, which is essential for the comparison to the experimental spectra. This has been demonstrated by us in the case of the floppy protonated peptide Ala-Ala-H⁺ in the gas phase [18,36]. A very good agreement of the calculated Car-Parrinello infrared spectrum of this molecule with the IR-MPD (InfraRed Multi Photon Dissociation) experimental spectrum, obtained at a finite temperature of ~300 K, could be achieved because the simulated room-temperature dynamics of the gas phase peptide was able to take into account the continual conformational dynamics between the two major isomers of the molecule. This could not be achieved with standard static ab initio calculations [18,56]. The main thing is that we found that the most populated conformations of Ala-Ala-H⁺ were not the geometries at the bottom wells (identified with static calculations) but rather all conformations explored in going from the basin of one isomer to the basin of the other isomer.

ii) When calculating IR spectra from the dipole time correlation function, all anharmonic effects are naturally described. This is to be opposed to the two successive harmonic approximations usually adopted for the determination of IR spectra from static *ab initio* calculations (harmonic approximation of the potential energy surface at the optimised geometries and electrical harmonic approximation for the transition dipole moments). Both approximations are released in molecular dynamics, simply because they are not needed. In fact, the finite temperature dynamics takes place on all accessible parts of the potential energy surface, be they harmonic or anharmonic. The quality of the potential energy surface is entirely contained in the "ab-initio" force field used in Car-Parrinello dynamics, calculated at the DFT/BLYP level in our works. The good reproduction of the relative positions of the different active bands in our studies (gas phase [36] and liquid phase [24,34,35]) is a demonstration that this level of theory is correct. Moreover, the calculation of IR spectra with molecular dynamics is related only to the time-dependent dipole moment of the molecule, and it does not require any harmonic expansion of the transition dipole moments. Therefore, if the dipole moments and their fluctuations are accurately calculated along the trajectory, the resulting IR spectrum should be reliable too.

iii) One should also emphasize two more points showing that dynamics (in particular Car-Parrinello dynamics) goes beyond *ab initio* static calculations applied currently in the domain. The first point is connected with the calculation of the infrared spectrum of complex flexible molecules in the gas phase, for example large peptides and nucleic acids. These complex molecular systems display numerous energetically equivalent conformations which are difficult to be characterised with geometry optimisation searches. Dynamics does release this difficulty as the geometry at the minima on the potential energy surface is not needed. On the other hand, molecular dynamics is the natural method of statistical mechanics for modeling complex molecular systems immersed in liquids. This is obviously very important when considering biomimetic molecules which natural environment is liquid water. The characterisation of the vibrational finger-prints of these molecules should better be performed in the liquid phase if one wants to get a proper structural information to be compared with the natural environment of the molecule. Inclusion of a few water molecules hydrogen bonded to the main hydrophilic sites of the molecule is nowadays commonly performed in *ab initio* geometry optimisations and harmonic frequency calculations, though still challenging. Properly characterising the surrounding hydration shells is nonetheless out of reach of static calculations. Calculations proposed by some authors (see for instance refs. [57–59]) with up to tens of water molecules clustered around the solute are still not convincing as these clusters do not possess any temperature. Water clusters surrounding a biomimetic solute at 0 K can not mimic the 'true' solvation shell of the solute at finite 300 K temperature. There are in particular hydrogen bonds between the solute and water displayed in solute-water clusters at 0 K that will only exist because there is no temperature in the molecular system, or because the full solvation shell (i.e. two or three surrounding shells) is not taken into account in the calculation. This has been shown by us in refs. [34,59] in the case of the solvated uracil.

3 Generalities on Car-Parrinello molecular dynamics simulations

In *ab initio* Car-Parrinello molecular dynamics (CPMD) [21], the nuclei are treated classically and the electrons quantum mechanically within the DFT formalism. CPMD simulations consist in solving Newton's equations of motion at finite temperature, with the forces that act on the nuclei deriving from the Kohn-Sham energy. Plane-waves and pseudo-potentials are used to represent valence electrons. Contrary to Born-Oppenheimer MD where the Schrödinger equation for the electronic configuration of the system is solved at each time step of the dynamics (i.e. at each new configuration of the nuclei), the Schrödinger equation is solved only once at the beginning of CPMD simulations. The wave-function is subsequently propagated in time adiabatically with the nuclei propagation. This is denoted "the fictitious dynamics of the electrons", and relies on the use of a fictitious electron mass. Typical fictitious mass values are comprised between 100 and 500 a.u. associated with time steps of 1-5 a.u. A detailed presentation of the

Car-Parrinello molecular dynamics method can be found in refs. [60,61].

The DFT-based Car-Parrinello simulations performed in our work follow the general set-up of *ab initio* molecular dynamics simulations that can be found in our previous publications [34,35,62,63]. We use the Becke, Lee, Yang and Parr (BLYP) gradient-corrected functional [64,65] for the exchange and correlation terms. The one-electron orbitals are expanded in a plane-wave basis set with a kinetic energy cut-off of 70 Ry restricted to the Γ point of the Brillouin zone. Medium soft norm-conserving pseudopotentials of the Martins-Trouillier type [66] are used. The core-valence interaction of C, N and O is treated by *s* and *p* potentials with pseudization radii of 1.23, 1.12 and 1.05 *au*, respectively (taking the same radius for *s* and *p*), while H atoms are treated as a *s* potential with a 0.5 *au* radius. We should mention that the 70 Ry energy cut-off is mostly determined by the convergence of the H and O pseudopotentials. Energy expectations are calculated in reciprocal space using the Kleinman-Bylander transformation [67]. Simulations were performed at constant volume using a fictitious electron mass of 500 *au*, a time step of 5 *au* (0.12 femtoseconds). Liquid simulations have been performed within periodic boundary conditions. Starting configurations have been prepared using classical force field simulations, where we chose the precise number of water molecules in the simulation cell; this is done in order to fix the correct density of such a small simulated liquid system. We refer the reader to Ref. [34,63] for more details on the procedure. Our liquid phase simulations take into account ~ 50 water molecules surrounding the solute molecule (N-methylacetamide presented in the present work), i.e. two complete hydration shells and part of the third one. For gas phase simulations, we refer the reader to ref. [36] for details on the choice of the cubic box length, and the use of the decoupling technique of Martyna and Tuckerman [68] in order to eliminate the effect of the periodic images of the charge density.

The dipole moment of the box cell is calculated with the Berry phase representation, as implemented in the Car-Parrinello framework [69]. Briefly, in the limit where the Γ point approximation applies, the electronic contribution to the cell dipole moment \mathbf{M}_α^{el} (where $\alpha = x, y, z$) is given by [70]:

$$\mathbf{M}_\alpha^{el} = \frac{e}{|\mathbf{G}_\alpha|} \Im \ln z_N \quad (9)$$

where $\Im \ln z_N$ is the imaginary part of the logarithm of the dimensionless complex number $z_N = \langle \Psi | e^{-i\mathbf{G}_\alpha \cdot \hat{\mathbf{R}}} | \Psi \rangle$, \mathbf{G}_α is a reciprocal lattice vector of the simple cubic supercell of length L ($\mathbf{G}_1 = 2\pi/L(1,0,0)$, $\mathbf{G}_2 = 2\pi/L(0,1,0)$, $\mathbf{G}_3 = 2\pi/L(0,0,1)$) and $\hat{\mathbf{R}} = \sum_{i=1}^N \hat{\mathbf{r}}_i$ denotes the collective position operator of the N electrons (or in other words the center of the electronic charge distribution). Ψ is the ground-state wave function. The quantity $\Im \ln z_N$ is the Berry phase, which in terms of a set of occupied Kohn-Sham orbitals $\psi_k(\mathbf{r})$ is computed as $\Im \ln z_N = 2\Im \ln \det \mathbf{S}$ with elements of the matrix \mathbf{S} given by $S_{kl} = \langle \psi_k | e^{-i\mathbf{G}_\alpha \cdot \hat{\mathbf{r}}} | \psi_l \rangle$ [70].

The IR spectra in our work are given as products $\alpha(\omega)n(\omega)$ expressed in cm^{-1} (decadic linear absorption coefficient) as a function of reciprocal wavenumber, ω , in cm^{-1} . The spectra have been smoothed with a window filtering applied in the time domain, i.e. each term of the correlation function $C(t)$ is multiplied by a gaussian function $\exp(-0.5\sigma(t/t_{max})^2)$, where t_{max} is the length of the simulation, and σ is 10 for gas phase simulations and 40 for liquid phase simulations.

When interested in the infrared spectrum of a solute immersed in liquid water, we face additional difficulties, namely separation of the absorption in solute and solvent contributions. Due to the limited statistics, the subtraction method used in experiment is not an option for simulation. The approach, initiated by Parrinello *et al.* [69] and that we first applied in the case of a solute immersed in liquid water in our calculation of the IR spectrum of aqueous uracil [34], is based on a decomposition of charge density using the Maximally localized Wannier Functions scheme of Marzari and Vanderbilt [71]. This enabled us to write the total electronic dipole moment $\mathbf{M} = \mathbf{m}_{solute} + \sum_W \mathbf{m}_W$ as the sum of the solute dipole moment (\mathbf{m}_{solute}) and the sum of the solvent dipole moments (\mathbf{m}_W). As a result, the total dipole correlation function $C_{MM}(t)$ needed in equation (8) ($C_{MM}(t) = \langle \mathbf{M}(t) \cdot \mathbf{M}(0) \rangle$) is resolved into the sum of three contributions $C_{W/W}(t)$, $C_{W/solute}(t)$ and $C_{solute/solute}(t)$ which are respectively the water-water dipole correlation function, the water-solute dipole correlation function and the solute-solute dipole correlation

function. The IR spectrum of the solute is calculated from only the Fourier transform of the self-correlation function $C_{solute/solute}(t)$. As discussed in Ref. [34], one disadvantage of this decomposition is that coupling with the surrounding solvent molecules is taken into account only through intermolecular polarization. The cross term $C_{W/solute}(t)$, which includes the remaining IR contributions arising from correlations between the motion of the solute and the solvent molecules, is ignored. In fact, this term is hard to compute as opposite positive and negative contributions make it difficult to reach convergence within the short simulation times of tens of picoseconds. As a last remark, while fully consistent in the sense that the molecular dipole moments rigorously sum to the cell dipole moment, the use of Wannier function for partitioning the electron density nonetheless introduces a certain arbitrariness for the definition of molecular dipole moments in solution.

An alternative approach to achieve a similar, but not necessarily equivalent, decomposition of the absorption signal is to return to the more fundamental current representation, involving the autocorrelation function of the total current, $\mathbf{j} = \dot{\mathbf{M}} = d\mathbf{M}/dt$:

$$n(\omega)\alpha(\omega) = \frac{2\pi}{3Vk_B T} \int_{-\infty}^{+\infty} \langle \mathbf{j}(0)\mathbf{j}(t) \rangle e^{i\omega t} dt. \quad (10)$$

Note again that in this formula we have taken into account a quantum correction factor of the form $\beta\hbar\omega/(1 - \exp(-\beta\hbar\omega))$. Since in the Born-Oppenheimer approximation the dipole moment of the system depends only on the atomic positions, one can make use of a chain rule to compute the current $\mathbf{j}(t)$ at time t :

$$j^\beta(t) = \sum_{i,\alpha} \frac{\partial M^\beta}{\partial x_i^\alpha}(t) \frac{dx_i^\alpha}{dt}(t) = \sum_{i,\alpha} \frac{\partial M^\beta}{\partial x_i^\alpha}(t) v_i^\alpha(t). \quad (11)$$

where \mathbf{M} is the total dipole moment of the system, \mathbf{x}_i is the position of atom i , and $\frac{\partial M^\beta}{\partial x_i^\alpha}$, $\alpha, \beta = x, y, z$, represent the components of the atomic polar tensor (APT) of atom i . The APT of atom i is thus defined as the derivatives of the dipole moment of the system with respect to infinitesimal displacements of atom i [72, 73]. This tensor is directly related to intensities of infrared bands in the double harmonic approximation [39], as we will show in section 6. Numerous calculations of APT have been performed in the solid phase or in the gas phase [72–75], and recently, a calculation of the APT of water molecules in liquid water has been performed [76].

In our applications in liquid water, the APT tensor was calculated for each atom in the solution using Density Functional Response Theory [73]. This was repeated every 1.2 fs over a ≈ 1 ps interval taken from the full trajectory in solution. In the gas phase calculations at low temperature, only one APT tensor of one conformation has been computed. We have used the Putrino *et al.* [31] implementation of the linear response to an applied electric field in the CPMD code [77]. In this scheme, the macroscopic polarization of the periodically replicated cell is defined using the Berry phase approach of Resta [70, 78]. The APT tensor for the whole system is then obtained from the Maxwell relation:

$$\frac{\partial M^\beta}{\partial x_i^\alpha} = \frac{\partial F_i^\alpha}{\partial \mathcal{E}^\beta} = \frac{\partial^2 E_{tot}}{\partial \mathcal{E}^\beta \partial x_i^\alpha} \quad (12)$$

where \mathcal{E} is an applied uniform electric field and \mathbf{F}_i the force acting on particle i . This allows for the calculation of the $3N$ elements of the system APT tensor through only three linear response calculations ($\beta = x, y, z$) [73].

In order to resolve the infrared spectrum of the whole system into solvent, solute and cross contributions, the total current of the system was separated into a solvent and a solute term:

$$j^\beta(t) = j_{water}^\beta(t) + j_{solute}^\beta(t) \quad (13)$$

by restricting the sum in eq. (11) to atoms of the solvent or the solute only. This natural decomposition of the current then provides a decomposition of the infrared spectrum which is different from the one obtained through the maximally localized Wannier orbitals. The infrared spectrum of the solute is now defined as

$$n(\omega)\alpha_{solute}(\omega) = \frac{2\pi}{3Vk_BT} \int_{-\infty}^{+\infty} \langle \mathbf{j}_{solute}(0) \mathbf{j}_{solute}(t) \rangle e^{i\omega t} dt. \quad (14)$$

A major argument in favor of the current scheme is that any partitioning of the charge density is thus avoided.

In our previous works we have shown that DFT-based Car-Parrinello molecular dynamics simulations (CPMD) yield very accurate infrared spectra of biomimetic molecules in the gas phase or immersed in aqueous solvent at room temperature, in terms of band-positions, band-shapes and band-intensities [34–36]. In particular, we have shown on the prototype gas phase protonated Ala-Ala-H⁺ peptide [36] that CPMD simulations are the proper tool to calculate IR absorption spectra of gas phase molecules undergoing multiple isomeric conformations at room temperature. This methodology has been applied to the IR spectra calculations of molecules such as Uracil [34], N-Methyl-Acetamide [35], and the protonated alanine peptide [36], either in liquid water or in the gas phase. We should mention here that in our applications we have systematically found that our calculated infrared spectra have to be blue-shifted by 100-120 cm⁻¹ so that the *whole* calculated bands can be aligned with their experimental counterparts. This holds true at least in the 800-2000 cm⁻¹ spectral region we have systematically investigated. Therefore, though our CPMD calculations do not give the proper absolute values of band positions, they do give the proper account of band-gaps between the different active bands. Among the effects leading to the very good agreement of our calculated infrared spectra with respect to the experiment (once the global translation has been applied) is the effective inclusion of the anharmonic effects in our finite temperature molecular simulations, unlike static ab initio calculations which are being performed in the harmonic approximation.

We stress again that a *single shift factor or global translation* is applied to the whole spectrum and not a scaling factor. This empirical finding is in contrast to static ab initio calculations where a scaling factor is used to correct the theoretical predictions with respect to the observed frequencies. The origin of this is at the moment unclear to us. Effects of the fictitious mass, which leads to instantaneous Car-Parrinello forces being different from Born-Oppenheimer ones whatever the fictitious mass value [79,80], is certainly important and indeed the blue-shift of 100-120 cm⁻¹ can be reduced when performing Car-Parrinello dynamics with smaller fictitious masses for the propagation of the electronic wave-function. This is though at the cost of more expensive simulations. However, it has been demonstrated in references [79,80] that the Car-Parrinello forces can be brought into good agreement with the BO forces by simply rescaling the ionic masses. Were it so, this would amount once more to a scaling factor and not a global translation. Such rescaling of the atomic masses should then be dependent on the vibrational mode considered, in a way not understood at the moment.

4 Assignment of vibrational modes

An accurate calculation of infrared spectra is one goal to achieve, the assignment of the active bands into individual atomic displacements or vibrational modes is another one, and this issue is essential to the understanding of the underlying molecular structural and dynamical properties. In molecular dynamics simulations, interpretation of the infrared active bands into individual atomic displacements is traditionally done using the vibrational density of states (VDOS) formalism. The VDOS is obtained by Fourier transformation of the atomic velocity auto-correlation functions:

$$VDOS(\omega) = \sum_{i=1,N} \int_{-\infty}^{\infty} \langle \mathbf{v}_i(t) \cdot \mathbf{v}_i(0) \rangle \exp(i\omega t) dt \quad (15)$$

where i runs over all atoms of the investigated system. There are no approximations in this formula. In addition, the VDOS can be decomposed according to each atom type in order to really get an interpretation of the vibrational bands in terms of individual atomic motions. This is done by restraining the sum over i in eq. (15) to the atoms of interest only. The advantage of the VDOS formalism is that all anharmonicities are taken into account in the calculation.

In the case of a harmonic molecular system though, velocities can be written as $\dot{q}_k = \dot{q}_k^0 \sin(\omega_k t + \phi_k)$, where \dot{q}_k^0 is the amplitude taken at the initial time, ω_k the wave-number of the k^{th} normal mode, and ϕ_k the phase, so that equation (15) can be re-expressed as

$$VDOS(\omega) = \sum_{i=1,N} \sum_{k,l} Z_{ik} Z_{il} \int_{-\infty}^{\infty} \langle \dot{q}_k(0) \dot{q}_l(t) \rangle e^{i\omega t} dt$$

where $\dot{\mathbf{x}}_i = \sum_k Z_{ik} \dot{q}_k$ is the transformation between cartesian and harmonic normal mode velocities, and $\langle \dot{q}_k(0) \dot{q}_l(t) \rangle$ is the correlation function between these harmonic velocities. Decoupling of harmonic modes leads to the following expression for the VDOS of a harmonic system:

$$VDOS(\omega) = \sum_{i=1,N} \sum_k Z_{ik}^2 \frac{\langle \dot{q}_k^0 \rangle}{2} \delta(\omega - \omega_k) \quad (16)$$

where $\langle \dot{q}_k^0 \rangle$ is taken as an ensemble average over the initial conditions of the dynamics. The VDOS of each individual atom (restraining the sum $\sum_{i=1,N}$ in equation 16 to one atom at a time) is therefore a sum of dirac functions which are localized at the wave-number of each normal mode ω_k , in the case of a strictly harmonic molecular system. This function is weighted by the amplitude Z_{ik}^2 which quantifies the participation of atom i into the normal mode k . Atoms that do not participate in a certain normal mode k thus make no contribution to the spectral signature of the VDOS.

It is mandatory to note here that VDOS spectra correspond to the vibrational density of states of the investigated molecular system, representing all its vibrational modes. However, only some of these modes will be infrared active or Raman active, so VDOS spectra can by no means substitute for IR or Raman spectra. They are used to interpret and assign vibrational modes which are active in IR or Raman spectra. Finally, when a solute is immersed into a solvent, the decomposition into individual atomic contributions of the VDOS can be interpreted as the contributions of the atoms of the solute on the one hand, and the contributions of the atoms of the solvent on the other hand. It is therefore easy to assess the couplings between solute and solvent atoms and their contributions to the VDOS spectral features of the solute molecule.

This methodology is illustrated in figure 1 in the case of the Car-Parrinello molecular dynamics simulation of N-Methyl-Acetamide in the gas phase at 20 K. We have reported the VDOS of the system as calculated through equation (15) at the bottom of the figure, and the individual VDOS associated with each atom type on top. We only report here the individual contributions of the four peptidic atoms (H, N, C, O) to the VDOS of N-Methyl-Acetamide. This is illustrated in the 1000-2000 cm^{-1} domain where the active bands can still be interpreted by simple combinations of stretching and bending movements of the atoms. Hence, one can note that the four atoms participate to the vibrational band located at $\sim 1600 \text{ cm}^{-1}$, while the oxygen atom does not contribute to the $\sim 1480 \text{ cm}^{-1}$ band. Again, the $\sim 1200 \text{ cm}^{-1}$ band can be decomposed to indicate the contributions of the C, N and H atoms, while the $\sim 1100 \text{ cm}^{-1}$ band displays contributions arising only from N and H atoms. One can thus infer that these bands are respectively related to C=O stretch and N-H bending (1600 cm^{-1}), N-H bending and N-C stretch (1480 cm^{-1}), N-H bending and N-C stretch (1200 cm^{-1}), and N-H bending (1100 cm^{-1}). Nonetheless, the amplitudes of the individual peaks can not be interpreted *per se*, so that we clearly can not conclude on the relative weight of each movement on each vibrational mode. Moreover, contributions arising from the methyl groups have to be considered in order to make a definitive statement on the assignment and interpretation of each

vibrational band. For a complete discussion on N-Methyl-Acetamide, see our ref. [35].

As can be seen from this example, analysis of the vibrational density of states is not so convenient and becomes rapidly fastidious as soon as the number of atoms in the molecular system increases. This analysis becomes basically unreachable for molecules that contain more than 10 atoms, N-Methyl-Acetamide being one of the more complex molecules on which this approach can be easily applied. One main bottleneck of this analysis is that VDOS signatures of each individual atom are spread all over the spectrum, which makes the interpretation and assignment of the vibrational bands relatively difficult. Last, but not least, VDOS do not provide a direct analysis in terms of stretches, bends and torsions, as we need for the interpretation of spectroscopic features of molecules. As we saw before, these motions can be inferred from simple molecules, even of the size of N-Methyl-Acetamide, but this is not true anymore when the size and complexity of the molecule increase. The same argument holds when we assign bands in the far-IR region where movements become more complicated, with couplings of torsional modes. Moreover, a quantitative analysis of the couplings between the movements that participate to the vibrations is out of reach with this approach. Another approach and formalism is required in order to answer to these questions.

Other methods have been proposed in the literature for assigning or approximating vibrational spectra, like spectral analysis [81–84], filtering [85–88], instantaneous normal mode analysis and its variants [89–92], or principal mode analysis (PMA) [93–96] and essential dynamics [97–99], both based on the diagonalisation of covariance matrices. Instantaneous normal mode analysis requires however many calculations of the Hessian system which can become quickly prohibitive for DFT-based MD simulation. Due to its simplicity, PMA has lately attracted attention [94, 95, 100–102].

Our new general definition for effective normal modes, that we will present in the following section, based on a localization principle in frequency space, not only leads at zero temperature to the usual normal mode analysis, but is also consistent with the PMA approach. Moreover, it provides another route to extract an effective Hessian and effective normal modes at finite temperature, without any computational cost.

5 New methodology for the assignment of vibrational modes: localized modes in terms of internal coordinates

For the purpose of assigning bands in vibrational spectra, we propose to construct modes q_k as (invertible) linear combinations of the natural coordinates that describe the dynamics of the systems:

$$\zeta_i(t) = Z_{ik} q_k(t) \Leftrightarrow q_k(t) = Z_{ki}^{-1} \zeta_i(t) \quad (17)$$

$$\dot{\zeta}_i(t) = Z_{ik} \dot{q}_k(t) \Leftrightarrow \dot{q}_k(t) = Z_{ki}^{-1} \dot{\zeta}_i(t). \quad (18)$$

In this equation and in the following, the Einstein summation convention will be used.

The coordinates $\zeta_i(t)$ can be either a set of cartesian coordinates $x_i(t)$ or a set of internal coordinates $S_j(t)$ or can also be mass weighted coordinates. As we saw above, since band assignment from VDOS is limited by the fact that these VDOS are delocalized in frequency we ask here that the mode q_k have VDOS as localized as possible in frequency. Noting $P_k^q(\omega)$ the power spectrum of mode k :

$$P_k^q(\omega) = \int_{-\infty}^{+\infty} \langle \dot{q}_k(0) \dot{q}_k(t) \rangle e^{i\omega t} dt \quad (19)$$

the modes are localized in frequency by minimizing the functional:

$$\Omega^{(n)} = \sum_k \left(\frac{\beta}{2\pi} \int_{-\infty}^{+\infty} d\omega |\omega^{2n}| P_k^q(\omega) - \left(\frac{\beta}{2\pi} \int_{-\infty}^{+\infty} d\omega |\omega^n| P_k^q(\omega) \right)^2 \right) \quad (20)$$

with respect to linear transformation \mathbf{Z} . This functional is parametrized by the free parameter n . In the case $n = 1$, the functional $\Omega^{(1)} = \sum_k (\langle \omega^2 \rangle_k - \langle \omega \rangle_k^2)$ is the sum of the spreads of the power spectra of modes k ; in the following we have rather made the choice $n = 2$.

This minimization criterion however is not sufficient for describing the localized modes q_k as the functional $\Omega^{(n)}$ can be minimized by setting $q_k = 0$. We thus require a normalisation and orthogonality criterion:

$$\frac{1}{2\pi} \int_{-\infty}^{+\infty} d\omega \left(\int_{-\infty}^{+\infty} \langle \dot{q}_k(0) \dot{q}_l(t) \rangle e^{i\omega t} dt \right) = \langle \dot{q}_k(0) \dot{q}_l(0) \rangle = k_B T \delta_{kl} \quad (21)$$

where T is the temperature and k_B the Boltzmann constant. This constraint then imposes the equipartition of the energy in the modes while keeping the modes decorrelated at equilibrium. This criterion will also ensure that the matrix \mathbf{Z} is invertible as required. The choice of introducing the temperature in this equation allows for a definition of the modes independant of T for a truly harmonic system.

To better understand the effect of band localization we can employ the Wiener-Khintchine theorem [103] to show a Cauchy-Schwartz relation:

$$P_{kl}^q(\omega) \leq \sqrt{P_{kk}^q(\omega) P_{ll}^q(\omega)} \quad (22)$$

where $P_{kl}^q(\omega)$ is a generalization of the power spectra (eq. 19) in a matrix form:

$$P_{kl}^q(\omega) = \int_{-\infty}^{+\infty} \langle \dot{q}_k(0) \dot{q}_l(t) \rangle e^{i\omega t} dt \quad (23)$$

If the power spectra $P_{kk}^q(\omega)$ and $P_{ll}^q(\omega)$ are well localized in frequency we can hope that for $k \neq l$ their overlap is very small and as a result the cross-correlation $P_{kl}^q(\omega)$ is very small for all ω such that the modes k and l have very small dynamical correlation. This is indeed expected for effective normal modes describing the dynamics of the system. Furthermore, we have seen in the preceeding section that in an harmonic system the normal mode possess delta like VDOS.

Minimization of the localization functional $\Omega^{(n)}$ does not require an iterative procedure like steepest-descent or conjugate gradients. It can be shown [37] that the solution of the minimization problem is solution of a much simpler linear system. The matrix \mathbf{Z} defining the localized modes from the choice of coordinates ζ_i is then solution of a generalized eigenvalue problem:

$$\mathbf{K}_{\zeta}^{(n)} \mathbf{Z}^{-1T} = \mathbf{K}_{\zeta}^{(0)} \mathbf{Z}^{-1T} \mathbf{\Lambda} \quad \text{with the constraint} \quad \mathbf{Z}^{-1} \mathbf{K}_{\zeta}^{(0)} \mathbf{Z}^{-1T} = \mathbf{I}_{3N} \quad (24)$$

where $\mathbf{\Lambda}$ is a diagonal matrix; the matrices $\mathbf{K}_{\zeta}^{(n)}$ are defined from the power spectra of the original ζ_i coordinates as:

$$\mathbf{K}_{\zeta}^{(n)} = \frac{\beta}{2\pi} \int_{-\infty}^{+\infty} d\omega |\omega|^n \mathbf{P}^{\zeta}(\omega) \quad (25)$$

where the matrix $\mathbf{P}^{\zeta}(\omega)$ is defined similarly as $\mathbf{P}^q(\omega)$ for the coordinates ζ_i and $\beta = \frac{1}{k_B T}$. The eigenvalues $\lambda_k^{(n)}$, diagonal elements of the matrix $\mathbf{\Lambda}$, are equal to:

$$\lambda_k^{(n)} = \int_{-\infty}^{+\infty} d\omega |\omega|^n P_{kk}^q \quad (26)$$

and are thus the averages $\langle \omega^n \rangle_k$ of ω^n over the power spectrum of mode k .

The definition of localized modes from their power spectra then appears to be equivalent to a generalized

eigenvalue problem with the same structure as normal mode analysis at zero temperature. In paper [37] we have employed cartesian coordinates to construct the localized normal modes and we have shown that the case $n = 2$ amounts to a generalization at finite temperature of the usual normal mode analysis where the Hessian at the optimized geometry is replaced by the average Hessian at finite temperature.

However, the use of cartesian coordinates may not be appropriate for large floppy molecules containing mobile groups like rotating CH_3 groups. This can be circumvented by using internal coordinates. To this aim we have taken the coordinates ζ_i as stretches, bends, torsions and out of plane angles. In order to build a complete set of non-redundant internal coordinates we have used the methods devised by Decius [104] or Pulay [19, 105]. Once these modes are defined, we can construct the Wilson \mathbf{B} matrix that describes the change of internal coordinates with respect to cartesian displacement for each configuration along the dynamics:

$$\mathbf{B} \equiv \frac{\partial S_j}{\partial x_i} \quad (27)$$

This \mathbf{B} matrix can then be used to evaluate the time-derivatives of the internal coordinates, needed for computing the power spectra matrices, from the cartesian velocities simply by applying

$$\dot{S}_j = \sum_i \frac{\partial S_j}{\partial x_i} \dot{x}_i \quad (28)$$

If we further need to compute accelerations of internal coordinates, numerical time-derivatives of the velocities have been used.

Figures 2 and 3 show the resulting VDOS of the effective normal modes obtained from first-principle molecular dynamics simulations for N-Methyl-Acetamide (NMA) in the gas phase and solvated in liquid water, respectively at low temperature and ambient temperature. For NMA, we have defined a complete set of non-redundant internal coordinates according to Pulay's definitions and localized modes were constructed as linear combinations of these coordinates as described above. In both cases, it can be seen that all modes are well localized in frequency, each showing mainly one peak with a spread usually less than 100 cm^{-1} . In the case of NMA in the gas phase, some modes have a nearly perfect localization indicating their harmonic character. In the liquid phase, overlap between bands can be large, as in the case of C-H stretch bands around 2900 cm^{-1} : these modes are however nearly decoupled from each other, and each mode still exhibits a high level of localization.

The method can be straightforwardly applied to larger systems and figure 4 shows the VDOS of effective normal modes calculated from the Car-Parrinello dynamics of a gas phase alanine dipeptide in the C7_{eq} conformation at 20 K. The first-principle molecular dynamics simulation of this molecule was done using the same general setup as for NMA. The initial set of non-redundant internal coordinates was also constructed following Pulay's prescriptions. At this low temperature, the system is mostly harmonic and the VDOS of the effective normal modes are all very well localized around each normal mode frequencies of this system, with very small or even no overlap between localized modes. This suggests that molecular dynamics simulations at low temperature may be an alternative route for determining normal modes, with the advantage that a perfect geometry optimization is not required contrary to the conventional hessian method. This is a net advantage for systems of increasing size and complexity.

Figure 5 illustrates the effect of the localization of VDOS in frequency in the case of the dynamics of the alanine di-peptide. At the top of the figure, VDOS of the two amide N-H groups of the alanine dipeptide calculated with the usual definition presented in equation (15) restricted to the N and H atoms only of each N-H group (see section 4) are shown, while the bottom of the figure illustrates the VDOS after the localization transformation (see section 5) using non-redundant internal coordinates as defined by Pulay [19]. In this latter case, the localized modes are identified as Amide III modes. Clearly, top of the figure exhibits peaks spread over the $1000\text{-}1800 \text{ cm}^{-1}$ range, in many different regions of the spectrum, and

are thus difficult to use as such to interpret the vibrational spectrum of the molecule. The two effective normal modes describing the two Amide III modes of alanine di-peptide are however perfectly localized around 1100 cm⁻¹. One can moreover note the ~20 cm⁻¹ frequency shift between both modes, indicating that the two N-H groups do not exhibit equivalent frequency signatures.

In order to describe the resulting effective normal modes in terms of internal motion, it is necessary to quantitatively determine the contribution of an internal coordinate in an effective normal mode. To this end, we employ a method derived from the Potential Energy Distribution as used in standard normal mode analysis [106–108]. PED is based on the contribution of each internal coordinates to the potential energy of an effective normal mode.

The eigenvectors \mathbf{Z} obtained after diagonalisation of the matrix $\mathbf{K}^{(2)}$ satisfy:

$$\mathbf{Z}^{-1} \mathbf{K}^{(2)} \mathbf{Z}^{-1T} = \mathbf{\Lambda} \Leftrightarrow \sum_{k,l} Z_{ik}^T K_{kl}^2 Z_{lj} = \lambda_{ij} \delta_{ij}$$

from which we deduce the normalization condition:

$$\frac{\sum_{kl} Z_{ki} K_{kl}^2 Z_{li}}{\lambda_{ii}} = 1$$

Then, the matrix P_{ij} that defines the contribution of the internal coordinate i to the normal mode j , can be expressed as [108]:

$$P_{ij} = \frac{\sum_l Z_{ji} K_{jl}^2 Z_{li}}{\lambda_i} \quad (29)$$

We have used here an alternative definition [107] of the PED as a slight variation of Morino's approach. Constructing the matrix $\mathbf{P}(k)$ indexed by the normal mode k as:

$$P_{ij}(k) = \frac{Z_{ik} Z_{jk} K_{ij}^2}{\lambda_k},$$

it has been recognised that the contribution of the internal coordinate i in this effective normal mode can be evaluated in two ways: either as a sum of the i^{th} row (or i^{th} column as the matrix is symmetric) - this is McCarthy's proposal - or as considering only the diagonal terms of $\mathbf{P}(k)$ which are necessarily positive:

$$P_{ii}(k) = P_{ik} = \frac{Z_{ik}^2 K_{ii}^2}{\lambda_k} \quad (30)$$

The resulting contributions need then be normalized and are generally expressed in percentage.

Table 1 shows the PED decomposition of some of the most important modes of NMA in the gas phase at 20 K, in both Cis and Trans conformations, and of Trans-Nma immersed in liquid water at 300 K. In the gas phase, three modes can be clearly identified as the three amide modes with Amide I being a nearly pure C=O stretch. In the liquid phase, this mode appears as a mixing of C=O stretch and N-H bend. Comparison of the gas phase and liquid phase for Trans-NMA shows that the C-N stretch and N-H bend are blue-shifted while the C=O stretch is red-shifted in the liquid phase. The resulting overlap of the C=O stretch and NH bend then leads to Amide modes which are mixed. These shifts are compatible with a picture where the zwitterionic form of the peptide bond is favored in solution, presumably due to the large dielectric constant of water and to hydrogen bonds formed between NMA and the surrounding water molecules. In this zwitterionic form, the CO bond acquires a more important single bond character, less stiff than in the neutral form, while the C-N bond acquires a double bond character. Table 1 also shows

that the other modes localized in this region of the spectrum mostly involve motion of the two terminal methyl groups displaying symmetric and anti-symmetric bends as defined by Pulay's non redundant internal coordinates [19].

Here again, concepts from the normal mode analysis are straightforwardly generalised to effective normal modes extracted from MD simulations and can thus help in the band assignments as well as in comparing results, at finite temperature, in the gas phase and condensed phase.

Another way of describing the calculated effective normal modes is simply by visual inspection of the atomic displacements associated with each mode. These displacements are defined as:

$$\delta x_i(k) = \frac{\partial x_i}{\partial q_k} \quad (31)$$

and can be obtained through the chain rule:

$$\frac{\partial x_i}{\partial q_k} = \sum_j \frac{\partial x_i}{\partial S_j} \frac{\partial S_j}{\partial q_k} \quad (32)$$

in which the matrix $\frac{\partial S_j}{\partial q_k}$ is simply the matrix \mathbf{Z} defining the effective normal modes. The first term, $\frac{\partial x_i}{\partial S_j}$ is obtained from Wilson's \mathbf{B} matrix defined as:

$$\mathbf{B} \equiv \frac{\partial S_j}{\partial x_i}. \quad (33)$$

This matrix however is $(3N - 6) \times 3N$ and is thus not invertible to obtain the needed $\frac{\partial x_i}{\partial S_j}$ terms. This can however be achieved by further requiring that the displacements $\delta x_i(k)$ satisfy the six Eckart conditions, i.e. no net translation nor rotation of the investigated molecular system. There also, we follow the general prescription of normal mode analysis in internal coordinates from Crawford and Fletcher [109] which results in:

$$\frac{\partial x_i}{\partial S_j} \equiv \mathbf{B}^- = \mathbf{M}^{-1} \mathbf{B}^T (\mathbf{B}^T \mathbf{M}^{-1} \mathbf{B})^{-1} \quad (34)$$

where \mathbf{B} is the Wilson's matrix and \mathbf{M} is the matrix of atomic masses. It is then possible to express the atomic displacement vector corresponding to mode k as:

$$\delta x_i(k) = (\mathbf{B}^- \mathbf{Z})_{ik} \quad (35)$$

Figure 6 displays some of the effective normal modes of NMA solvated in liquid water. On these figures, as on the PED reported in table 1, one can identify the participation of C=O stretch, NH bend and C-N stretch to the three Amide modes, while the two other modes displayed whose frequencies are in the same range as the Amide bands are deformations of the terminal methyl groups. See table 1 for the precise assignment of the modes.

We are thus able to extract effective normal modes from molecular dynamics simulations performed at finite temperature, either in the gas phase or for a solute immersed in the condensed phase. These modes are effective in the sense that they include temperature, all anharmonicities of the dynamics (potential energy surface anharmonicities, dipole anharmonicities), and dynamical isomerisation along the dynamics. From the effective normal mode localization, a PED analysis can be applied. These two steps are strictly similar to the analyses conducted in static ab initio calculations, with the net advantage of an underneath molecular dynamics simulation.

6 Infrared intensities

The localization of modes in frequency is very helpful for assigning observed bands in IR spectra. However, not all the computed modes may be observed in the spectra because their IR intensity is low. Sometimes many bands may be present in the same frequency domain and only a few may have significant IR intensity to be observable. Thus for the purpose of assigning IR spectra bands, it is of great interest to also have at disposal the IR intensity associated with each effective normal mode motion. This is the purpose of this paragraph.

Starting from the expression of the infrared spectrum as the Fourier-transform of the autocorrelation function of the dipole or its time derivative (current), see eq. (11), we can re-write the current in the newly introduced effective normal modes q_k as:

$$j^\beta(t) = \sum_k \frac{\partial M^\beta}{\partial q_k}(t) \frac{dq_k}{dt}(t) = \sum_k \frac{\partial M^\beta}{\partial q_k}(t) \dot{q}_k(t) \quad (36)$$

to re-express the infrared absorption as:

$$IR(\omega) = n(\omega)\alpha(\omega) = \frac{2\pi}{3Vk_B T} \int_{-\infty}^{+\infty} \sum_{k,l} \left\langle \left(\frac{\partial M^\beta}{\partial q_k} \dot{q}_k \right) (0) \left(\frac{\partial M^\beta}{\partial q_l} \dot{q}_l \right) (t) \right\rangle e^{i\omega t} dt \quad (37)$$

In this equation the velocity \dot{q}_k of the k^{th} effective normal mode is obtained as:

$$\dot{q}_k(t) = \frac{\partial q_k}{\partial S_t} \frac{\partial S_t}{\partial x_i} \dot{x}_i \quad (38)$$

from the internal coordinates definition. The APT tensors expressed in effective normal mode coordinates is a vector $\frac{\partial M^\beta}{\partial q_k}$ analogous to a time-dependent transition dipole vector. It can be calculated in a similar way using the chain rule:

$$\frac{\partial \mathbf{M}}{\partial q_k} = \frac{\partial \mathbf{M}}{\partial x'_i} \frac{\partial x'_i}{\partial S_t} \frac{\partial S_t}{\partial q_k} \quad (39)$$

where the first term, $\frac{\partial \mathbf{M}}{\partial x'_i}$ is the APT tensor in the laboratory frame, the second term is the generalized inverse of Wilson's matrix B which we have already introduced, while the third term is the matrix transform between effective normal modes and the internal coordinates. The first two terms are time-dependent and need to be calculated on the fly along the dynamics with an inversion of the B matrix for each configuration.

The first approximation to this otherwise exact formula for the infrared spectrum is to assume that the effective normal modes are decorrelated from each others, which gives:

$$IR(\omega) = n(\omega)\alpha(\omega) \approx \frac{2\pi}{3Vk_B T} \int_{-\infty}^{+\infty} \sum_k \left\langle \left(\frac{\partial M^\beta}{\partial q_k} \dot{q}_k \right) (0) \left(\frac{\partial M^\beta}{\partial q_k} \dot{q}_k \right) (t) \right\rangle e^{i\omega t} dt \quad (40)$$

as suggested by the definition of the effective normal modes that should minimize the overlap between their respective power spectra. In this approximation, the infrared spectrum is decomposed as a sum of effective normal modes contributions that can be computed in order to interpret the spectrum and assign observed IR bands. The rest of the paragraph is devoted to further approximations that allow to assign an intensity to each effective normal mode.

The next approximation to this sum of modes is to assume a kind of rotation-vibration decorrelation

assuming that the terms $\frac{\partial M^\beta}{\partial q_k}$ are decorrelated from the mode velocities, so that:

$$\left\langle \left(\frac{\partial M^\beta}{\partial q_k} \dot{q}_k \right) (0) \left(\frac{\partial M^\beta}{\partial q_k} \dot{q}_k \right) (t) \right\rangle = \left\langle \left(\frac{\partial M^\beta}{\partial q_k} \right) (0) \left(\frac{\partial M^\beta}{\partial q_k} \right) (t) \right\rangle \langle \dot{q}_k(0) \dot{q}_k(t) \rangle \quad (41)$$

and

$$IR(\omega) = n(\omega) \alpha(\omega) \approx \frac{2\pi}{3Vk_B T} \sum_k \int_{-\infty}^{+\infty} \left\langle \left(\frac{\partial M^\beta}{\partial q_k} \right) (0) \left(\frac{\partial M^\beta}{\partial q_k} \right) (t) \right\rangle \langle \dot{q}_k(0) \dot{q}_k(t) \rangle e^{i\omega t} dt \quad (42)$$

Noting that $\partial \mathbf{M} / \partial q_k$ is a vector which we can write as a norm times a unit vector: $\partial \mathbf{M} / \partial q_k = |\partial \mathbf{M} / \partial q_k| \mathbf{u}(t)$, and further assuming decorrelation between the two, we can write:

$$IR(\omega) = \frac{2\pi}{3Vk_B T} \sum_k \int_{-\infty}^{+\infty} \left\langle \left| \frac{\partial \mathbf{M}}{\partial q_k} (0) \right| \cdot \left| \frac{\partial \mathbf{M}}{\partial q_k} (t) \right| \right\rangle \langle \mathbf{u}_k(0) \cdot \mathbf{u}_k(t) \rangle \langle \dot{q}_k(0) \dot{q}_k(t) \rangle e^{i\omega t} dt \quad (43)$$

which is our expression for the IR spectrum including mode decoupling and complete rotation-vibration decorrelation.

In a final approximation, we can assume that the term $\langle \dot{q}_k(0) \dot{q}_k(t) \rangle$ describing the modes dynamics varies much more rapidly than the two other terms, which thus can be evaluated at time $t = 0$ only. The infrared spectrum is therefore simply written as:

$$IR(\omega) = \frac{2\pi}{3Vk_B T} \sum_k A_k \int_{-\infty}^{+\infty} \langle \dot{q}_k(0) \dot{q}_k(t) \rangle e^{i\omega t} dt \quad (44)$$

where the A_k 's are the mode intensities:

$$A_k = \left\langle \left| \frac{\partial \mathbf{M}}{\partial q_k} (0) \right|^2 \right\rangle. \quad (45)$$

In this form, the infrared spectrum is a weighted sum of the effective normal modes power spectra, whose intensities are obtained from the Atomic Polar Tensor or transition dipole vectors in a similar way as in the double harmonic approximation. In the examples below, the intensities A_k are obtained using the APT calculated on one single configuration only of the system instead of a true configuration average $\langle \left| \frac{\partial \mathbf{M}}{\partial q_k} (0) \right|^2 \rangle$, this allows for a rapid estimation of band intensities for the purpose of assigning bands.

This is however not necessary in our framework, as upon the change of coordinates made to introduce the effective normal modes, all information about the system dynamics is still retained at this stage and the IR spectrum can still be calculated exactly as in eq. (37) or with a minimal approximation, eq. (40) that still allows for a decomposition of the total spectrum in a sum of effective modes contributions.

In the following, we will be presenting calculations of IR intensities following equation (42) with the APT tensor and correlation calculated over one pico-second of the dynamics, or using equation (44) in which only one APT tensor is calculated for the first configuration of the dynamics.

6.1 Applications

Comparisons between the infrared spectrum calculated through the Fourier transform of the dipole correlation function (equation (8)) and the Fourier transform of the localized normal modes (equation (44)) making use of APT tensors is presented in figures 7 and 8. Comparisons are presented for the Trans-NMA molecule in the gas phase (Car-Parrinello dynamics at 20 K) and immersed in liquid water (Car-Parrinello molecular dynamics simulation at 300 K). The whole spectrum is presented in the gas phase, while only

the 1000-2000 cm^{-1} domain is presented in the liquid phase. In both cases, it can be seen that the infrared spectrum of Trans-NMA calculated through equation (44) captures the main features of the exact infrared spectrum calculated through equation (8). Moreover, the result is not sensitive to the choice of the APT tensor as can be concluded from the condensed phase calculation in figure 8, be it sampled over part of the dynamics (here one pico-second) or fixed from the first configuration of the dynamics. In particular, structural deformations taken into account in the APT tensors calculated over one pico-second of the trajectory do not give substantial further informations to the infrared intensities and band widths. In the gas phase, both spectra are similar, with mainly an underestimation of the intensity of the Amide I band at $\sim 1600 \text{ cm}^{-1}$ and of the C-H stretching band above 2000 cm^{-1} when equation (44) is applied. The N-H stretching band above 2000 cm^{-1} is overestimated. In the liquid phase calculation, one can see that the three main Amide bands are present in all calculations, though the Amide II band located at $\sim 1400 \text{ cm}^{-1}$ is merged with the Amide I band in the calculations done through the APT tensors. The disappearance of the Amide II band in the APT calculation of the IR spectrum could be due to the choice of definition of the molecular current from the APT tensor, which gives a completely different decomposition of the density in comparison to the Wannier orbital decomposition. Note here (see our paper [35]) that the Car-Parrinello spectrum, calculated with the Wannier orbital decomposition, gives an IR spectrum of NMA immersed in water which is in very good agreement with the experiment.

Figure 9 reports the 1000-1800 cm^{-1} frequency domain of the calculated infrared spectrum of Trans-NMA obtained from our Car-Parrinello trajectory at 20 K (dashed black line) and its decomposition into its fundamental normal mode contributions as described with our formalism of the IR intensities. The intensities of the bands are reported in table 1. The band located at $\sim 1595 \text{ cm}^{-1}$ clearly comes solely from the Amide I mode (stretching C=O as from 95%, red line). The broad band roughly located between 1300 and 1500 cm^{-1} can be nicely decomposed into three contributions. The 1440 cm^{-1} shoulder on the upper frequency part is due to the Amide II mode (bending N-H as from 42%, N-Terminal methyl symmetric stretching as from 17% and stretching C-N as from 10%, blue line); two modes take part in the central part of the broad band, respectively coming from the C-Terminal asymmetric methyl C-H bendings (85%, purple line) and the N-Terminal symmetric methyl C-H bendings (77%, black line); the lower frequency part is due to the C-Terminal symmetric methyl C-H bendings (95%, orange line). Features arising from the symmetric and asymmetric C-H bendings can thus be nicely separated over the whole infrared spectrum, as well as the contributions arising from each of the terminal methyl groups. Finally, the $\sim 1180 \text{ cm}^{-1}$ band results from the Amide III mode, which is a composition of C-N stretching (32%), N-H bending (30%) and C=O stretching (12%) in our calculation.

The same 1000-1800 cm^{-1} frequency domain decomposition is presented in figure 10 in the case of the calculated infrared spectrum of Trans-NMA from our Car-Parrinello molecular dynamics trajectory in aqueous water at 300 K (black dashed line). The intensities of the bands are reported in table 1 and the transition dipole moments are shown as red arrows in figure 6. As can be easily seen, the $1400\text{-}1600 \text{ cm}^{-1}$ active band is composed of the Amide I mode (now interpreted as 54% of C=O stretching and 38% of N-H bending) on its upper frequency part and of the Amide II mode (interpreted as 48% N-H bending, 24% C-N stretching and 13% N-Terminal C-H symmetric stretching) on its lower part. Both modes thus merge into a broad band in the liquid phase, while both modes are now composed of more mixing between C=O and N-H motions. The active band located at $\sim 1230 \text{ cm}^{-1}$ is only composed of the Amide III mode (interpreted as 38% stretching C-N, 35% bending N-H and 12% stretching C=O, which is very similar to the gas phase interpretation). The band located at $\sim 1350 \text{ cm}^{-1}$ is a merging of the methyl symmetric C-H bending motions, the C-Terminal methyl being at the upper frequency part and the N-Terminal methyl at the lower frequency part.

Another illustration is presented in figure 11 in the case of the gas phase Car-Parrinello trajectory of the octa-alanine peptide chain at $\sim 20 \text{ K}$. As can be seen again (top of figure 11), the approximated infrared spectrum (calculated through equation (42)) is very close to the exact infrared spectrum (calculated through equation (8)), the biggest discrepancy arising again from the intensity of the C-H stretching modes at $\sim 2800 \text{ cm}^{-1}$. The three main bands obtained in the 1000-1800 cm^{-1} frequency domain can be nicely decomposed into individual contributions as proposed at the bottom of figure 11. The band at $\sim 1570 \text{ cm}^{-1}$ is composed of the Amide I C=O stretching modes, arising from the different carbonyl

groups along the octa-peptide chain, the band at $\sim 1420\text{ cm}^{-1}$ results from the Amide II N-H bending modes of the different amide groups along the octa-peptide chain, while the $\sim 1130\text{ cm}^{-1}$ band comes from the Amide III modes. For each Amide band, one can see that each decomposed sub-band is peaked at a different absolute position value, the sum of which giving rise to the broadening of the resulting final band.

7 Discussion–Conclusions

We have reviewed here Car-Parrinello molecular dynamics simulations for the calculation of infrared spectra of complex molecular systems, either in the gas phase or in the condensed phase, like here immersed in a liquid environment, at finite temperature. Although first-principle simulations have proved powerful for this aim, interpretation of the resulting vibrational spectrum is often not straightforward, based on the study of VDOS for selected coordinates as illustrated here in the case of N-Methyl-Acetamide. We have here presented a new formalism in order to extract localized effective normal modes from the trajectories in order to interpret and assign the active vibrational bands of the IR spectra. This method is quite general and can be applied in gas, solid or liquid phase. The effective normal modes thus constructed can be written as linear combinations of linear coordinates, being more adapted to solid state or stiff molecules [37], and we have introduced here effective normal modes as linear combinations of internal coordinates. This is better adapted to situations where part of the system/molecule is floppy.

Our methodology has been illustrated on the N-Methyl-Acetamide (NMA) molecule, on the alanine dipeptide molecule and on the more complex octa-alanine peptide chain. The effective normal modes were constructed here in terms of individual internal coordinates (taken as non-redundant as defined by Pulay [19]). We have shown that our formalism is able to localize effective vibrational modes in the frequency domain, allowing for a clear separation of the different modes.

The construction of the effective normal modes is formally only a change of variables and no information is lost from the dynamics. Only the representation of the dynamics has changed, allowing for an easier band assignment since, in this representation, the VDOS are as localized as possible; they still have however a width and possibly complex shape reflecting the complex dynamics in the system. The modes thus extracted are effective in the sense that they include temperature, all anharmonicities of the dynamics (potential energy surface anharmonicities and dipole anharmonicities), and a complex dynamics such as dynamical isomerisation along the time. We have previously shown [37] that this method is close to PMA for a choice $n = -2$ and cartesian coordinates, and is a generalization at finite temperature of the usual normal mode analysis for the choice $n = 2$. Here, we have also shown that concepts like PED can be straightforwardly adapted to our approach. We are then able to compute PED contributions in order to give a proper and precise assignment of each effective mode as illustrated for the gas phase and solvated NMA.

In this paper, we also introduced infrared intensities calculated for the effective normal modes. These are very helpful indicators for the assignment of bands in the observed or calculated IR spectra. Different ways of calculating the IR intensities associated with each effective mode have been devised and we have shown that they correspond in each case to an approximation of the IR 'exact' spectrum. In the simplest case, it amounts to computing the APT tensor in the new set of effective normal mode coordinates, defining a transition moment similar to the normal mode analysis case first introduced by Pulay [19]. As can be seen in the illustrations on the peptide models, the reconstructed IR spectra hence computed from the band intensity and band VDOS compare very well to the exact ones calculated through the Fourier Transform of the time correlation of the dipole moment.

These intensities, although representing the IR spectra in a simplified way as described above, can be also useful in providing an approximation to the IR spectrum when a full calculation may be difficult or costly (calculation of molecular dipoles by orbital rotation or computation of currents). This should be particularly true for Raman spectra for which the approach presented here can bring a useful approximation at a minimal numerical cost.

The effective normal mode method discussed here can be seen as fitting the best harmonic model to the computed dynamics, as we have previously discussed in cartesian coordinates [37]. This suggests that

the method could be used to reparameterize intramolecular classical force-fields to better represent the molecular vibrations in a given environment, including anharmonicity in an effective way. It is however also expected that mode localization will not be very efficient in strong anharmonic situations, as in the case of a complex molecular dynamics where isomerisation between different isomeric conformations occur. In such situations, the methodology discussed here can be slightly modified in order to extract the influence of each isomeric conformation on the effective normal modes computed from the entire dynamics. The velocity correlations can indeed be weighted by the amount of time spent by the molecule in each of the isomeric basins, and the localization principle can still be applied for vibrations in each basin. This is current work in progress to assess the validity of this approach. Another envisioned development for anharmonicity is the inclusion of higher order terms in the definition of effective modes (e.g. a second order polynomial).

Finally, one of the biggest advantages of first-principle molecular dynamics for the computation of vibrational spectra is the explicit description of the electronic system. This allows, in principle, the prediction of other type of vibrational spectroscopy than infrared spectra, like Raman spectroscopy or vibrational circular dichroism.

Acknowledgments. MPG acknowledges support from Genopole-France through the program 'ATIGE' Action Thématique Incitative de Génopole. The authors thank CINES (Montpellier, France) and IDRIS (Orsay, France) for generous access to their computational facilities.

References

- [1] S. Krimm and J. Bandekar. *Adv. Prot. Chem.*, 38:181, 1986.
- [2] H. Torii and M. Tasumi. *J. Chem. Phys.*, 96:3379, 1992.
- [3] *Biological Applications of Raman Spectroscopy, Vol I.* edited by T. Spiro, Wiley-Interscience, New-York, 1987.
- [4] *Infrared Spectroscopy of Biomolecules.* edited by H. Mantsch and D. Chapman, Wiley-Liss, New-York, 1996.
- [5] R. Huang, J. Kubelka, W. Barber-Armstrong, R. Silva, S.M. Decatur, and T.A. Keiderling. *J. Am. Chem. Soc.*, 126:2346, 2004.
- [6] P. Bour and T.A. Keiderling. *J. Phys. Chem. B.*, 109:5348, 2005.
- [7] J.R. Zheng JR and M.D. Fayer. *J. Am. Chem. Soc.*, 129:4328, 2007.
- [8] I.J. Finkelstein, J.R. Zheng, H. Ishikawa, S. Kim, K. Kwak, and M.D. Fayer. *Phys.Chem.Chem.Phys.*, 9:1533, 2007.
- [9] S. Bagchi S, Y.S. Kim, A.K. Charnley, A.B. Smith, and R.M. Hochstrasser. *J. Phys. Chem. B.*, 111:3010, 2007.
- [10] S. Gnanakaran and R.M. Hochstrasser. *J. Am. Chem. Soc.*, 123:12886, 2001.
- [11] P. Hamm, M. Lim, and R.M. Hochstrasser. *J. Phys. Chem. B.*, 102:6123, 1998.
- [12] S. Woutersen and P. Hamm. *J. Phys. Chem. B.*, 104:11316, 2000.
- [13] P.C. Chen and C.C. Joyner. *J. Mol. Struct.*, 799:23, 2006.
- [14] P.C. Chen and C.C. Joyner. *J. Phys. Chem. A.*, 110:7989, 2006.
- [15] J. Oomens, G. Meijer, and G.V. Helden. *J. Phys. Chem. A.*, 105:8302, 2001.
- [16] J. Lemaire, P. Boissel, M. Heninger, G. Maucclair, H. Mestdagh, A. Simon, S. Le Caer, J.M. Ortega, F. Glotin, and P. Maitre. *Phys. Rev. Lett.*, 89:273002-1, 2002. and references therein.
- [17] J. Oomens, B.G. Sartakov, G. Meijer, and G. Von Helden. *Int. J. Mass Spectrom.*, 254:1, 2006.
- [18] G. Grégoire, M.P. Gaigeot, D.C. Marinica, J. Lemaire, J.P. Schermann, and C. Desfrancois. *Phys. Chem. Chem. Phys.* Accepted for publication in *Phys. Chem. Chem. Phys.*, DOI: 10.1039/b618094a.
- [19] P. Pulay, G. Fogarasi, F. Pang, and J.E. Boggs. *J. Am.Chem.Soc.*, 101:2550, 1979.
- [20] K.K. Irikura, R.D. Johnson, and R.N. Kacker. *J. Phys. Chem. A.*, 109:8430, 2005.
- [21] R. Car and M. Parrinello. *Phys. Rev. Lett.*, 55:2471, 1985.
- [22] P.L. Silvestrelli and M. Parrinello. *J. Chem. Phys.*, 111:3572, 1999.
- [23] Vincent Dubois, Paolo Umari, and Alfredo Pasquarello. *Chem. Phys. Lett.*, 390(1-3):193-198, 2004.
- [24] Jan-Willem Handgraaf, Evert Jan Meijer, and Marie-Pierre Gaigeot. *J. Chem. Phys.*, 121(20):10111-10119, 2004.
- [25] M.P. Gaigeot, R. Vuilleumier, M. Sprik, and D. Borgis. *J. Chem. Theory Comput.*, 1:772, 2005.
- [26] M.-P. Gaigeot and M. Sprik. *J. Phys. Chem. B*, 108:7458, 2004.
- [27] M.-P. Gaigeot and M. Sprik. *J. Phys. Chem. B*, 107:10344, 2003.
- [28] Radu Iftimie and Mark E. Tuckerman. *J. Chem. Phys.*, 122(21):214508, 2005.
- [29] M. Nonella, G. Mathias, and P. Tavan. *J. Phys. Chem. A*, 107:8638, 2003.
- [30] Aneta Jezierska, Jaroslaw J. Panek, Aleksander Koll, and Janez Mavri. *J. Chem. Phys.*, 126:205101, 2007.
- [31] Anna Putrino, Daniel Sebastiani, and Michele Parrinello. *J. Chem. Phys.*, 113:7102, 2000.
- [32] A. Putrino and M. Parrinello. *Phys. Rev. Lett.*, 88:176401, 2002.
- [33] A. Miani, S. Raugai, P. Carloni, and M.S. Helfand. *J. Phys. Chem. B.*, 111:2621, 2007.
- [34] M.-P. Gaigeot and M. Sprik. *J. Phys. Chem. B.*, 107:10344, 2003.
- [35] M.-P. Gaigeot, R. Vuilleumier, M. Sprik, and D. Borgis. *J. Chem. Theor. Comput.*, 1:772, 2005.
- [36] C. Marinica, G. Grégoire, C. Desfrancois, J.P. Schermann, D. Borgis, and M.-P. Gaigeot. *J. Phys. Chem. A.*, 110:8802, 2006.
- [37] M. Martinez, M.P. Gaigeot, D. Borgis, and R. Vuilleumier. *J. Chem. Phys.*, 125:144106, 2006.
- [38] D.A. McQuarrie. *Statistical Mechanics*. Harper-Collins Publishers: New York, 1976.
- [39] E.B. Wilson, J.C. Decius, and P.C. Cross. *Molecular Vibrations*. McGraw-Hill, New-York, 1955.
- [40] G. Zerbi. Introduction to the theory of vibrational frequencies and vibrational intensities. In *Vibrational intensities in infrared and Raman spectroscopy*, by W. B. Person and G. Zerbi. Elsevier Scientific Publishing Company, Amsterdam - Oxford - New York, 1982.
- [41] L. Pauling and E.B. Wilson. *Introduction to quantum mechanics, with applications to chemistry*. Dover, 1985.
- [42] R. Kubo, M. Toda, and N. Hashitsume. *Statistical Physics*, volume II. Springer Verlag, second edition, 1991.
- [43] R. Iftimie and M.E. Tuckerman. Decomposing total IR spectra of aqueous systems into solute and solvent contributions: A computational approach using maximally localized Wannier orbitals. *J. Chem. Phys.*, 122:214508, 2005.
- [44] H. Ahlborn, B. Space, and P.B. Moore. *J. Chem. Phys.*, 112:8083, 2000.
- [45] J. Borysow, M. Moraldi, and L. Frommhold. *Mol. Phys.*, 56:913, 1985.
- [46] R. Ramirez, T. Lopez-Ciudad, P. Kumar, and D. Marx. *J. Chem. Phys.*, 121:3973, 2004.
- [47] D. Levesque, J.J. Weis, P. Marteau, J. Obriot, and F. Fondere. *Mol. Phys.*, 54:1161, 1985.
- [48] Y. Guissani, D. Levesque, J.J. Weis, and D.W. Oxtoby. *J. Chem. Phys.*, 77:2153, 1982.
- [49] T. Tassaing, Y. Danten, M. Besnard, E. Zoidis, J. Yarwood, Y. Guissani, and B. Guillot. *Mol. Phys.*, 84:769, 1995.
- [50] B. Guillot. *J. Chem. Phys.*, 95:1543, 1991.
- [51] R. Vuilleumier and D. Borgis. *J. Chem. Phys.*, 111:4251, 1999.
- [52] A. Staib and D. Borgis. *Chem. Phys. Lett.*, 271:232, 1997.
- [53] W.J. Glover and P.A. Madden. *J. Chem. Phys.*, 121:7293, 2004.
- [54] E.A. Pavlatou, P.A. Madden, and M. Wilson. *J. Chem. Phys.*, 107:10446, 1997.
- [55] P.A. Madden and R.W. Impey. *Chem. Phys. Lett.*, 123:502, 1986.
- [56] B. Lucas, G. Grégoire, J. Lemaire, P. Maitre, J.M. Ortega, A. Rupenyan, B. Reimann, J.P. Schermann, and C. Desfrancois. *Phys.Chem.Chem.Phys.*, 6:2659, 2004.
- [57] V.I. Danilov, T. van Mourik, and V.I. Poltev. *Chem. Phys. Lett.*, 429:255, 2006.
- [58] O.V. Shishkin, L. Gorb, and J. Leszczynski. *J. Phys. Chem. B.*, 104:5357, 2000.
- [59] M.P. Gaigeot and M. Ghomi. *J. Phys. Chem. B.*, 105:5007, 2001.
- [60] D. Marx, M. E. Tuckerman, and M. Parrinello. *J. Phys.: Condens. Matter A*, 12:153-159, 2000.
- [61] R. Vuilleumier. Density functional theory based ab initio molecular dynamics using the car-parrinello approach. In *Computer Simulations in Condensed Matter Systems: From Materials to Chemical Biology Volume 1*, by Mauro Ferrario, Kurt Binder, and Giovanni Ciccotti. Springer, Berlin, 2006.
- [62] M.-P. Gaigeot and M. Sprik. *J. Phys. Chem. B.*, 108:7458, 2004.
- [63] R. Vuilleumier and M. Sprik. *J. Chem. Phys.*, 115:3454, 2001.
- [64] A. Becke. *Phys. Rev. A.*, 38:3098, 1988.
- [65] C. Lee, W. Yang, and R.G. Parr. *Phys. Rev. B.*, 37:785, 1988.
- [66] N. Trouillier and J.L. Martins. *Phys. Rev. B.*, 43:1993, 1991.

- [67] L. Kleinman and D.M. Bylander. *Phys. Rev. Lett.*, 48:1425, 1982.
- [68] G.J. Martyna and M.E. Tuckerman. *J. Chem. Phys.*, 110:2810, 1999.
- [69] M. Bernasconi, P.L. Silvestrelli, and M. Parrinello. *Phys. Rev. Lett.*, 81:1235, 1998.
- [70] R. Resta. *Phys. Rev. Lett.*, 80:1800, 1998.
- [71] N. Marzari and D. Vanderbilt. *Phys. Rev. B*, 56:12847–12865, 1997.
- [72] J. Cioslowski. *J. Am. Chem. Soc.*, 111:8333, 1989.
- [73] X. Gonze. *Phys. Rev. B*, 55:10337, 1997.
- [74] F. De Proft, J. M. L. Martin, and P. Geerlings. *Chem. Phys. Lett.*, 250:393, 1996.
- [75] A. Pasquarello and R. Car. *Phys. Rev. Lett.*, 79:1766, 1997.
- [76] A. Pasquarello and R. Resta. *Phys. Rev. B*, 68:174302, 2003.
- [77] J. Hutter et al. *CPMD, version 3.7.1*. IBM Research Division, IBM Corp and Max Planck Institute Stuttgart.
- [78] P. L. Silvestrelli and M. Parrinello. *Phys. Rev. Lett.*, 82:3308, 1999.
- [79] P. Tangney. *J. Chem. Phys.*, 124:044111, 2006.
- [80] P. Tangney and S. Scandolo. *J. Chem. Phys.*, 116:14, 2002.
- [81] J. M. Bowman, X. Zhang, and A. Brown. *J. Chem. Phys.*, 119:646, 2003.
- [82] M. Kaledin, A. Brown, A. Kaledin, and J. M. Bowman. *J. Chem. Phys.*, 121:5646, 2004.
- [83] J. Kohanoff. *Comp. Mat. Sc.*, 2:221, 1994.
- [84] G. Onida, W. Andreoni, J. Kohanoff, and M. Parrinello. *Chem. Phys. Lett.*, 219:1, 1994.
- [85] Pnina Dauber-Osguthorpe, David J. Osguthorpe, Peter S. Stern, and John Moulton. *J. of Comput. Phys.*, 151:169, 1999.
- [86] Richard B. Sessions, David J. Osguthorpe, and Pnina Dauber-Osguthorpe. *J. Phys. Chem.*, 99:9034, 1995.
- [87] Pnina Dauber-Osguthorpe and David J. Osguthorpe. *J. Comput. Chem.*, 14:1259, 1993.
- [88] Pnina Dauber-Osguthorpe and David J. Osguthorpe. *J. Am. Chem. Soc.*, 112:7921, 1990.
- [89] R. M. Stratt. *Acc. Chem. Res.*, 28:201, 1995.
- [90] M. Buchner, B. Ladanyi, and R. M. Stratt. *J. Chem. Phys.*, 97:8522, 1992.
- [91] M. Cho, G. R. Fleming, S. Saito, I. Ohmine, and R. M. Stratt. *J. Chem. Phys.*, 100:6672, 1994.
- [92] R. E. Larsen and R. M. Stratt. *J. Chem. Phys.*, 110:1036, 1999.
- [93] Ronald M. Levy, Olivia de la Luz Rojas, and Richard A. Friesner. *J. Phys. Chem.*, 88:4233, 1984.
- [94] R. A. Wheeler, H. Dong, and S. E. Boesch. *Chem. Phys. Chem.*, 4:382, 2003.
- [95] R. A. Wheeler and H. Dong. *Chem. Phys. Chem.*, 4:1227, 2003.
- [96] B. R. Brooks, D. Janezic, and M. Karplus. *J. Comp. Chem.*, 16:1522, 1995.
- [97] A. Amadei, A. B. M. Linssen, and H. J. C. Berendsen. *Proteins: Struct. Funct. Genet.*, 17:412, 1993.
- [98] F. Tama, F.X. Gadea, O. Marques, and Y.H. Sanejouand. *Proteins: Struc. Func. Genet.*, 41:1, 2000.
- [99] F. Tama and Y.H. Sanejouand. *Protein Eng.*, 14:1, 2001.
- [100] M. Schmitz and P. Tavan. *J. Chem. Phys.*, 121(24):12247–12258, 2004.
- [101] M. Schmitz and P. Tavan. *J. Chem. Phys.*, 121(24):12233–12246, 2004.
- [102] A. Strachan. *J. Chem. Phys.*, 120:1, 2004.
- [103] R. Kubo, M. Toda, and N. Hashitsume. *Statistical Physics II-nonequilibrium statistical mechanics*. Springer, Heidelberg, 2 edition, 1985.
- [104] J.C. Decius. *J. Chem. Phys.*, 17:1315, 1949.
- [105] P. Pulay and F. Fogarasi. *J. Chem. Phys.*, 96:2856, 1992.
- [106] Y. Morino and K. Kuchitsu. *J. Chem. Phys.*, 20:1809, 1952.
- [107] J.C. Whitmer. *J. Mol. Spec.*, 68(2):326, November 1977.
- [108] W.J. McCarthy, L. Lapinski, M. J. Nowak, and L. Adamowicz. *J. Chem. Phys.*, 108:10116, 1998.
- [109] B.L. Crawford and W.H. Fletcher. *J. Chem. Phys.*, 19:141, 1951.

Mode	Freq. (cm^{-1})	Intens. (km/mol)	Contributions from PED
gas phase cis-NMA			
Amide III	1255	58.9	C1 sym.bend.:47%, st.C2-N:22%, b.N-H:7%
C1 sym.bend	1322	217.5	C1 sym.bend.:98%
Amide II	1365	0.9	b.N-H:42 %,st.C=O:32%
Méthyl def.	1396	22.3	C1 asym.bend.:61%, C3 sym.bend.:28%
Méthyl def.2	1413	20.7	C1 asym.bend.:46%, C3 asym.bend.:39%
Amide I	1598	451.4	st C=O: 92%, b. N-H: 4%
gas phase trans-NMA			
Amide III	1179	63.7	st.C2-N:32%, b.N-H:30%, st.C=O:12%
C1 sym.bend.	1320	33.0	C1 sym.bend.:95%, C1 asym.bend.:3%
C3 sym.bend.	1363	36.7	C3 sym.bend.:77%, b.N-H:11%
Méthyl bend.	1394	40.2	C1 asym.bend.:85%, b.N-H:4%
Amide II	1420	37.6	b.N-H:42%, C3 sym.bend.:17%, st.C2-N:10%
Amide I	1596	330.7	st.C=O:95%, b.N-H:2%
solvated trans-NMA			
Amide III	1243	124.8	st.C2-N:38%, b.N-H:35%, st.C=O:12%,
C1 sym. def.	1349	47.3	C1 sym.bend.:93%
C3 sym. def.	1374	99.2	C3 sym.bend.:79%, b.N-H:6%
Amide II	1493	478.9	b.N-H:48%, st.C2-N:24%, C3 sym.bend.:13%
Amide I	1515	235.9	st.C=O:54%, b.N-H:38%

Table 1. PED and IR intensities of localized modes between 1100 and 1700cm^{-1} for gas phase Cis- and Trans-NMA and solvated Trans-NMA, extracted from first-principle Molecular Dynamics simulations at 20 K (gas phase) and 300 K (liquid phase). Effective modes are defined using Pulay's non redundant internal coordinates [19]. st: stretch, b: bend, sym bend: symmetric bend, asym bend: asymmetric bend.

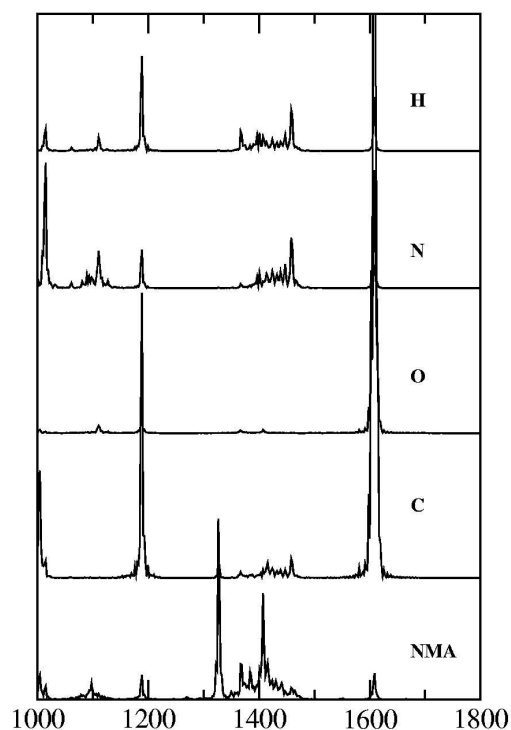


Figure 1. Vibrational density of states (VDOS) of isolated Trans-NMA (N-Methyl-Acetamide) obtained in a Car-Parrinello molecular dynamics at 20 K. The total VDOS is displayed at the bottom of the figures and the various contributions from individual atoms of the peptidic group (H, N, C, O) are given above. C and O respectively stand for the carbonyl carbon and oxygen atom of NMA, N and H for the amide nitrogen and hydrogen atoms.

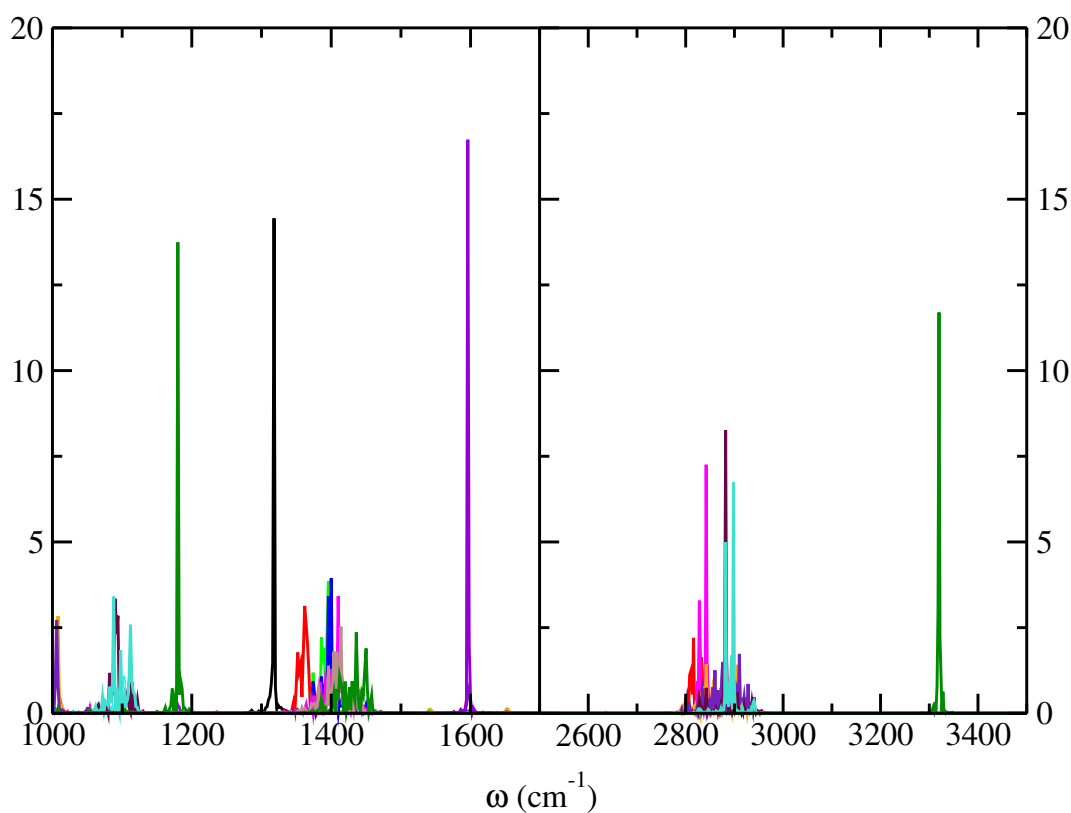


Figure 2. Power spectra of the localized modes for gas phase Trans-N-Methyl-Acetamide at 20 K.

The effective normal modes are computed with the formalism described in section 5 using non-redundant internal coordinates as defined by Pulay [19].

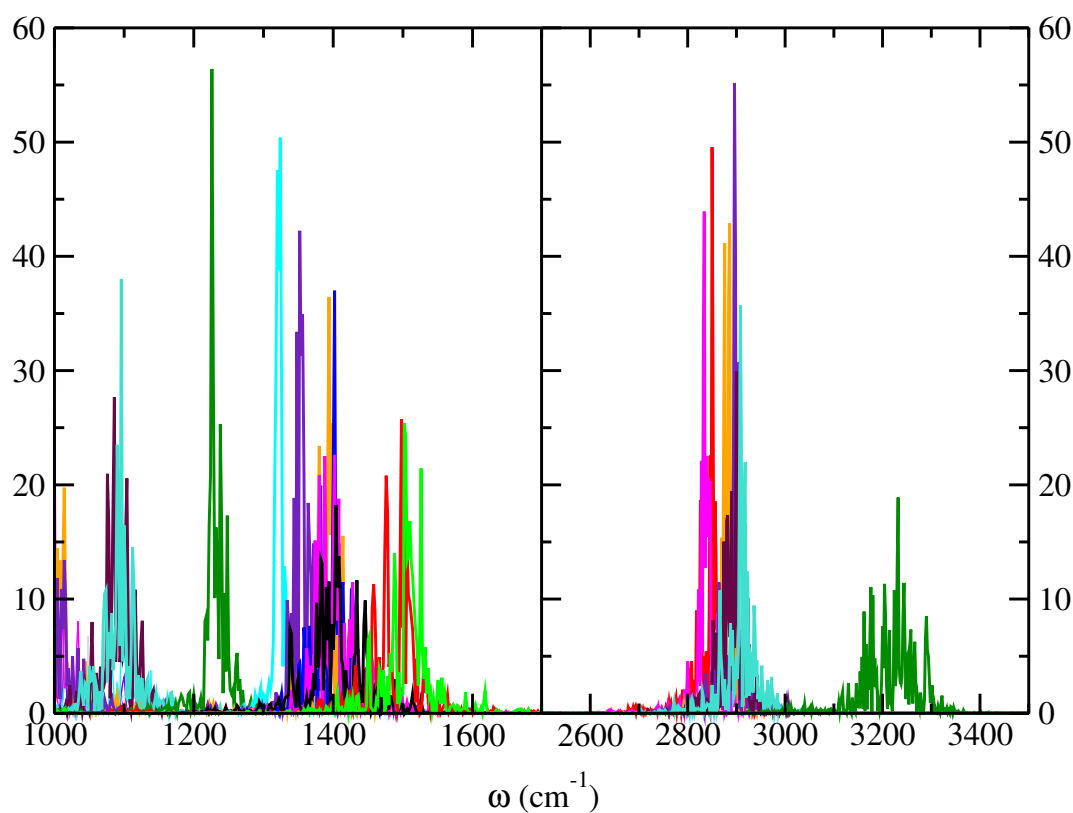


Figure 3. Power spectra of the localized modes for solvated Trans-N-Methyl-Acetamide at 300 K.

The effective normal modes are computed with the formalism described in section 5 using non-redundant internal coordinates as defined by Pulay [19].

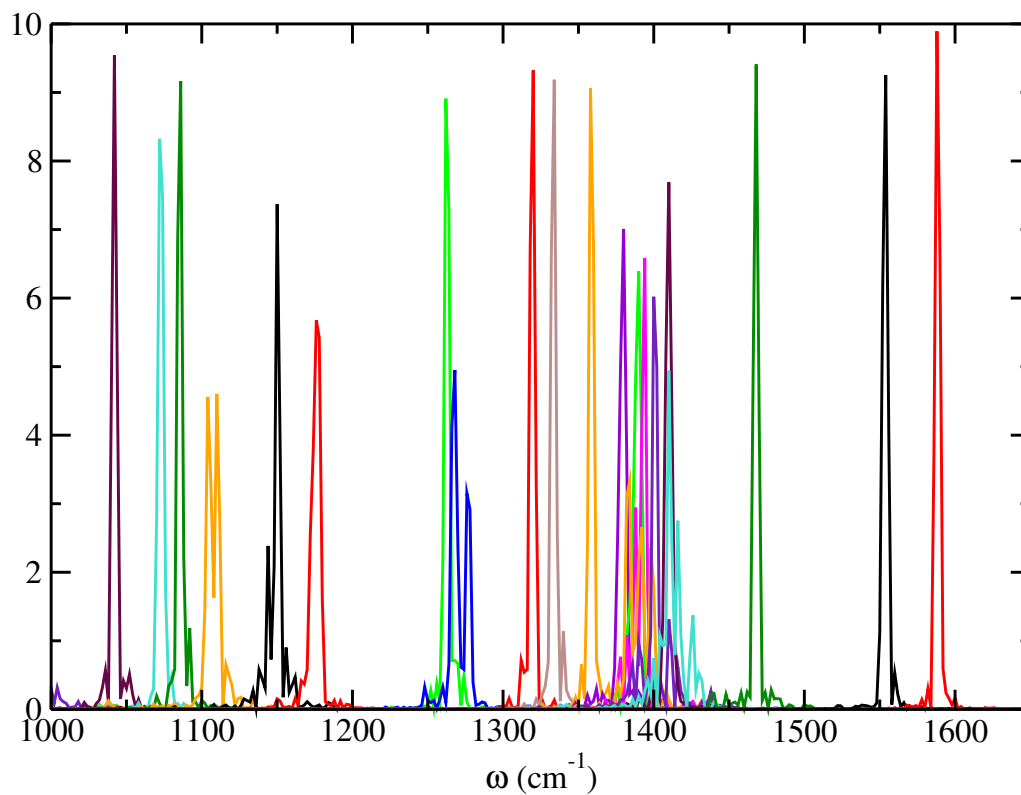


Figure 4. Power spectra of the localized modes for gas phase alanine di-peptide in the $C7_{eq}$ conformation, from the gas phase Car-Parrinello dynamics at 20 K. The effective normal modes are computed with the formalism described in section 5 using non-redundant internal coordinates as defined by Pulay [19].

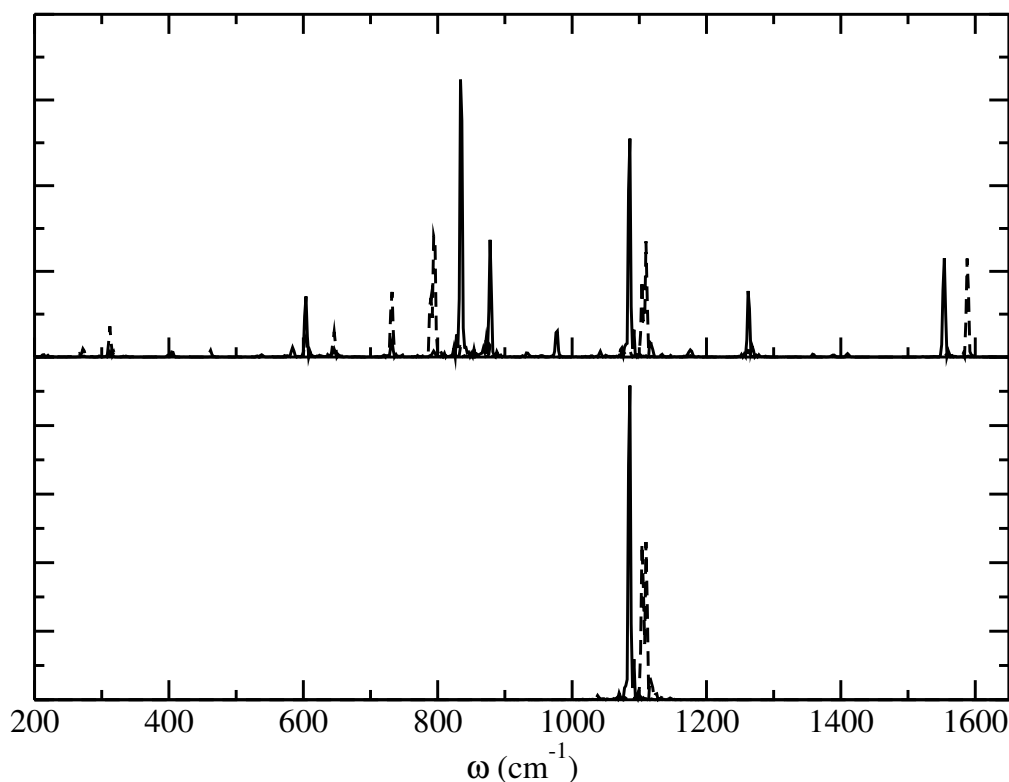


Figure 5. Top: VDOS of the two amide N-H groups of an alanine dipeptide calculated with the usual definition in equation (15) restricted to the N and H atoms only of each N-H group, see section 4. One N-H group is identified with the black line, the other with the dashed black line. See text for the related discussion. Bottom: Power spectra of the effective normal modes associated to the N-H Amide III modes after the localization transformation presented in section 5 using non-redundant internal coordinates as defined by Pulay [19]. See text for the related discussion.

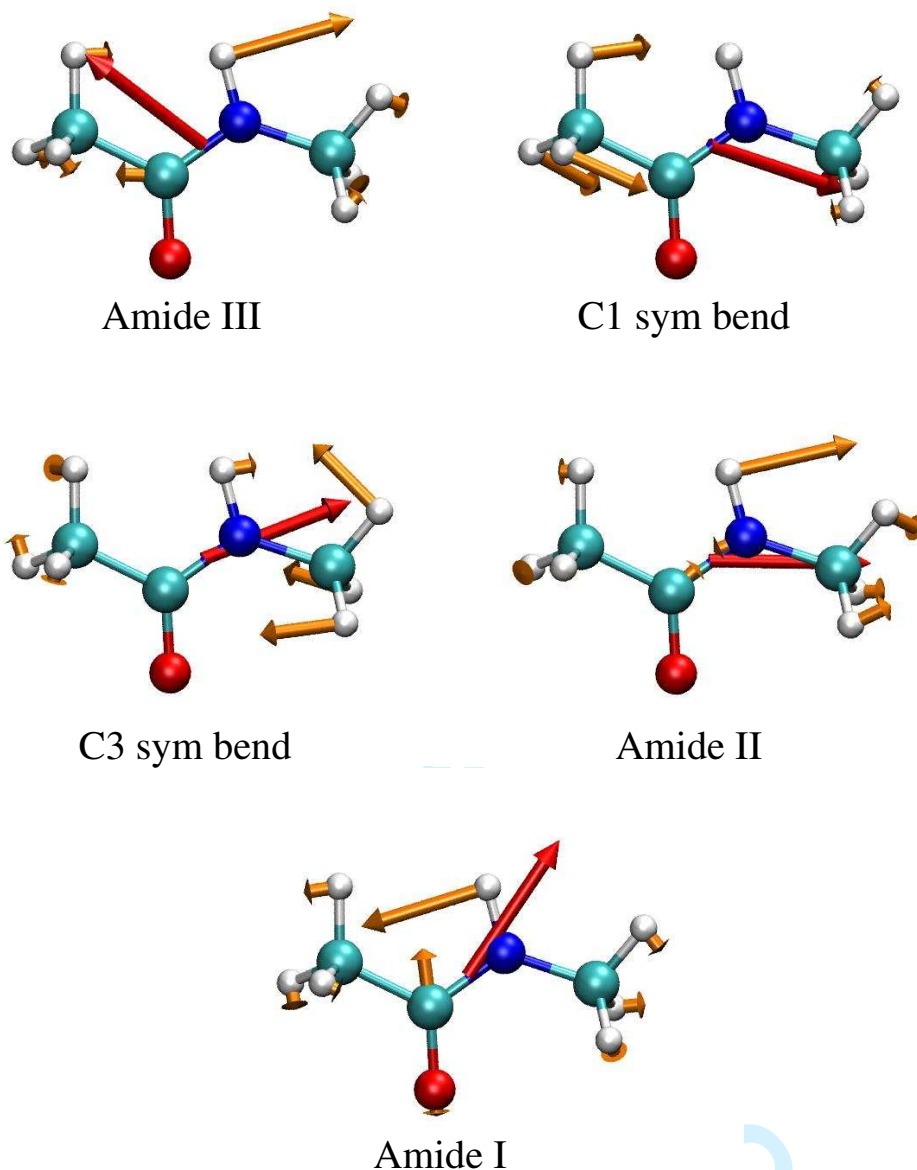


Figure 6. Atomic displacements (orange arrows) for some selected effective normal modes of solvated Trans-NMA. For each mode, the transition dipole as defined by eq. 39 is also shown as a red arrow.

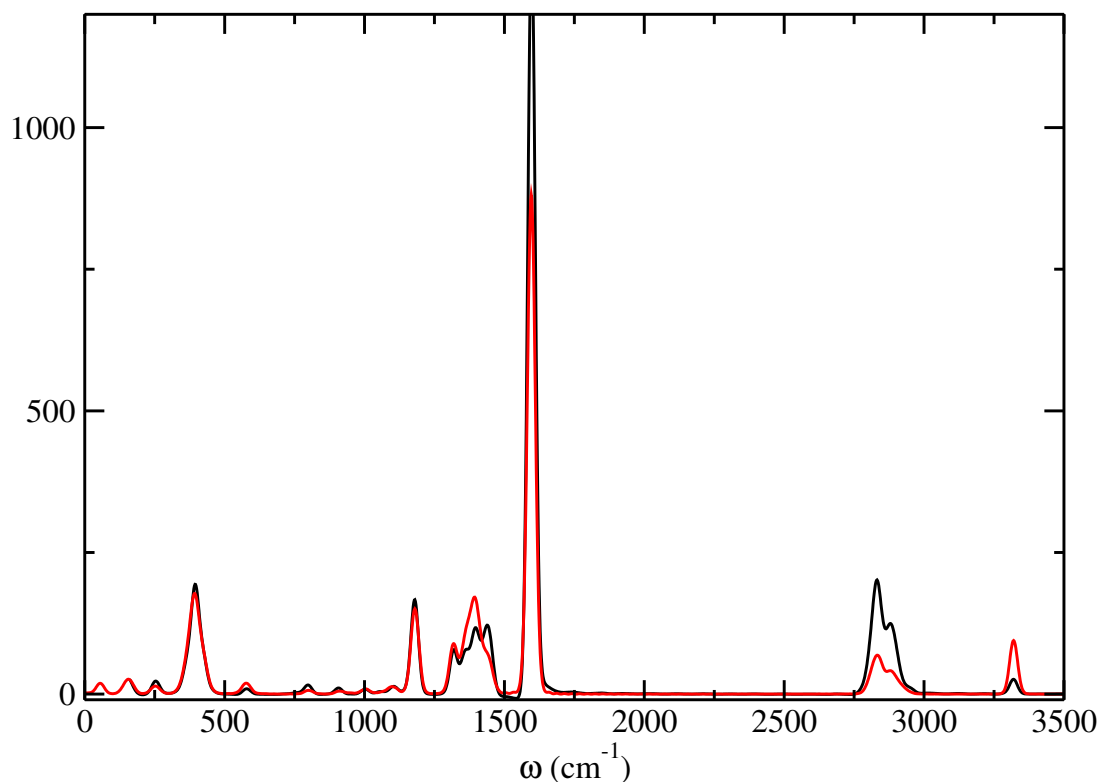


Figure 7. Comparison between the infrared spectrum calculated through the Fourier transform of the dipole correlation function (equation (8), black line) and the Fourier transform of the localized normal modes (equation (44), red line) making use of APT tensors (APT tensor calculated at the first configuration of the dynamics). Comparison presented for the Car-Parrinello molecular dynamics of Trans-NMA in the gas phase at 20 K.

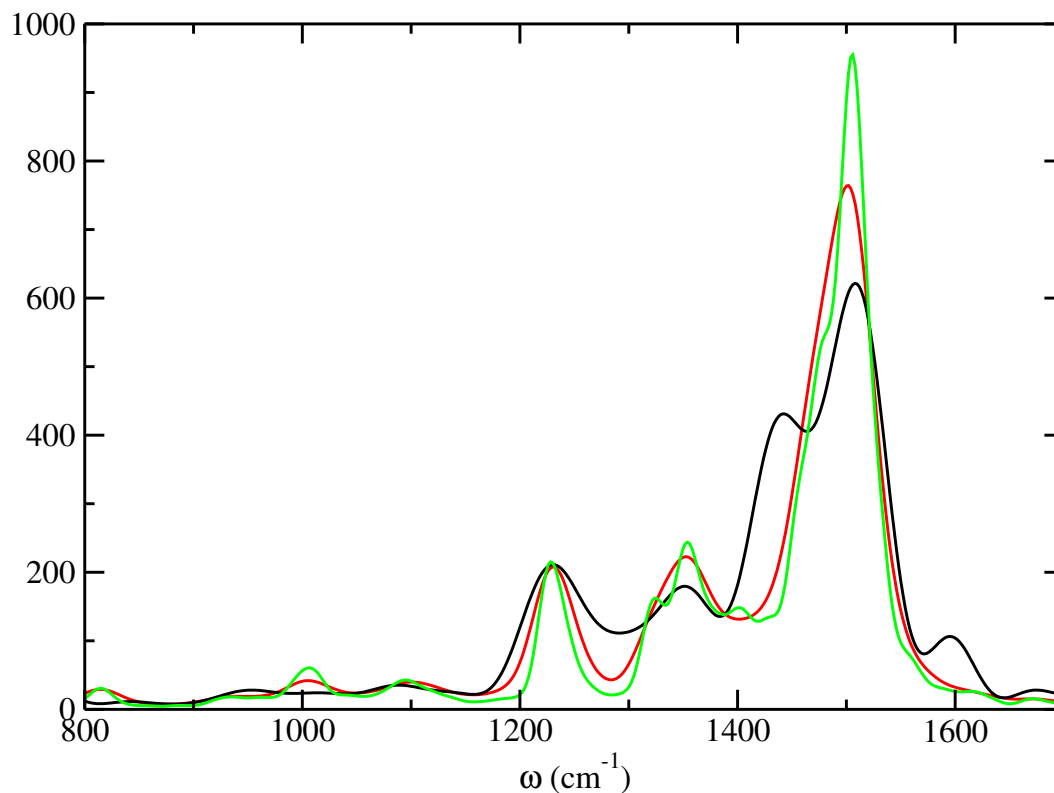


Figure 8. Comparison between the infrared spectrum calculated through the Fourier transform of the dipole correlation function (equation (8), black line) and the Fourier transform of the localized normal modes (equation (44), red and green lines) making use of APT tensors (red: APT tensor sampled over one pico-second trajectory, green: APT tensor calculated at the first configuration of the dynamics). Comparison presented for the Car-Parrinello molecular dynamics of Trans-NMA immersed in liquid water at 300 K.

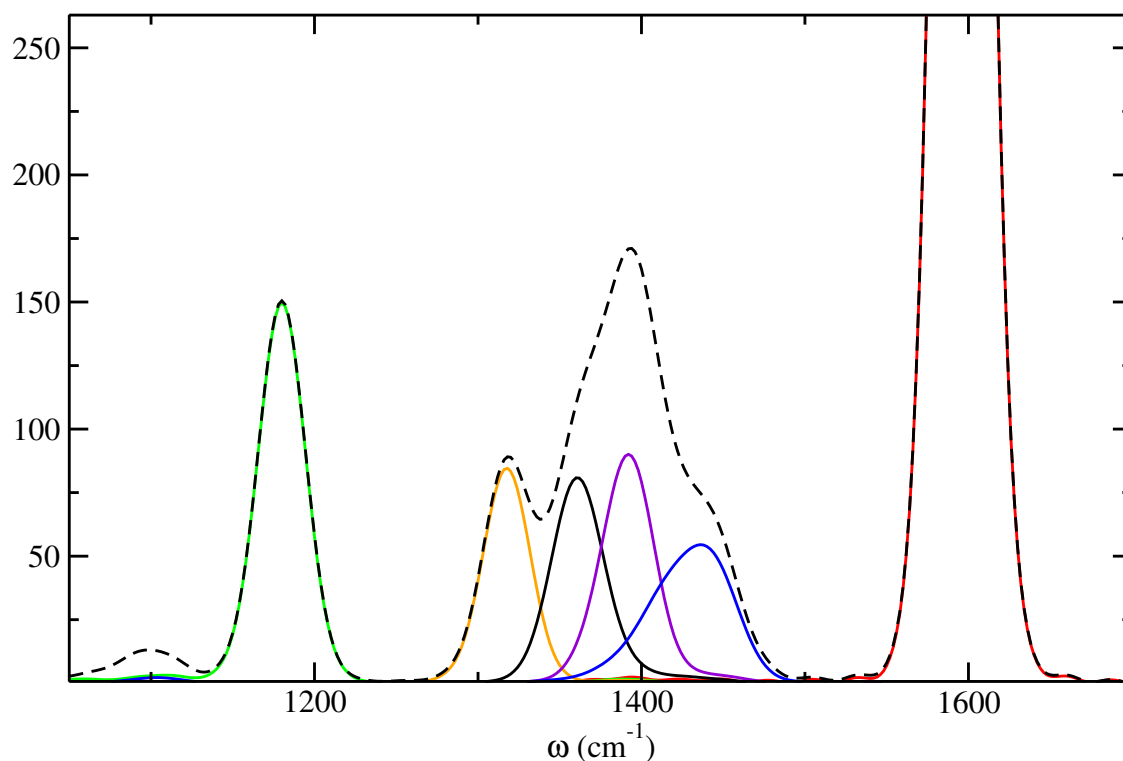


Figure 9. Decomposition of the infrared spectrum intensity of Trans-NMA from the 20 K simulation in the gas phase. Decomposition into the different contributions arising from the Amide normal modes in the 1000-1800 cm^{-1} frequency domain: Dashed Black: total IR spectrum; Red: Amide I mode; Blue: Amide II mode; Orange, black and purple: methyl groups C-H bending; green: Amide III mode.

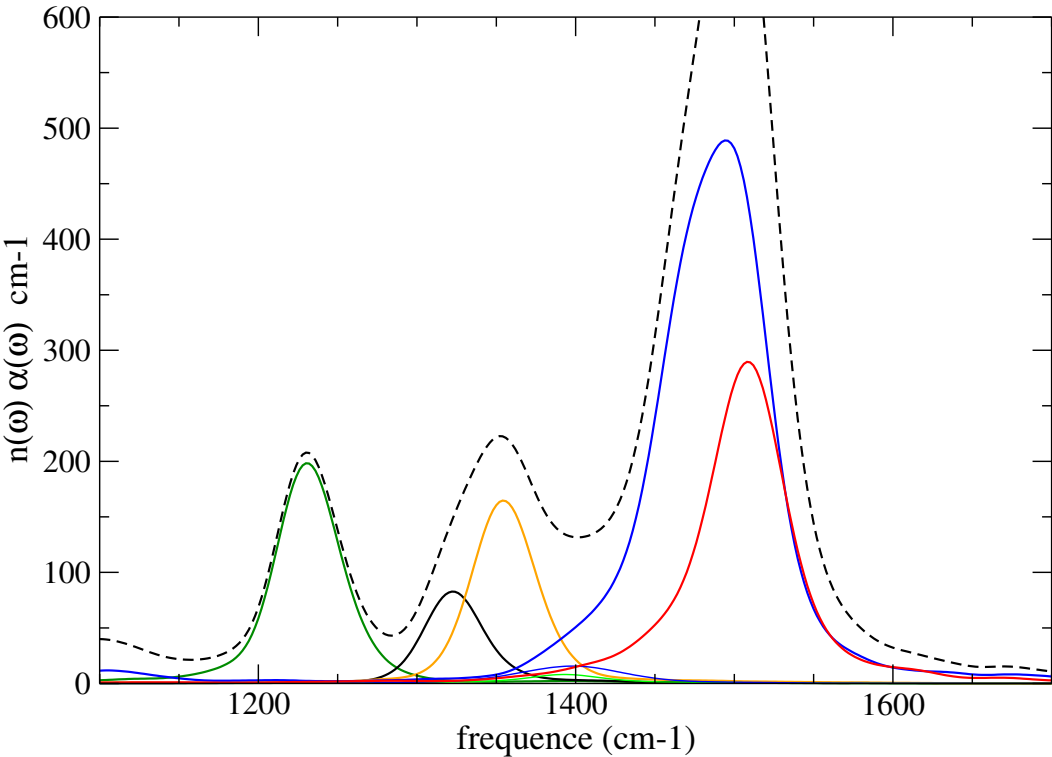


Figure 10. Decomposition of the infrared spectrum intensity of Trans-NMA solute from the 300 K aqueous Trans-NMA simulation. Decomposition into the different contributions arising from the Amide normal modes in the 1000-1800 cm^{-1} frequency domain: Dashed Black: total IR spectrum; Red: Amide I mode; Blue: Amide II mode; Orange and black: methyl groups C-H bending modes, respectively from the C-Terminal side and N-Terminal side of the molecule; green: Amide III mode.

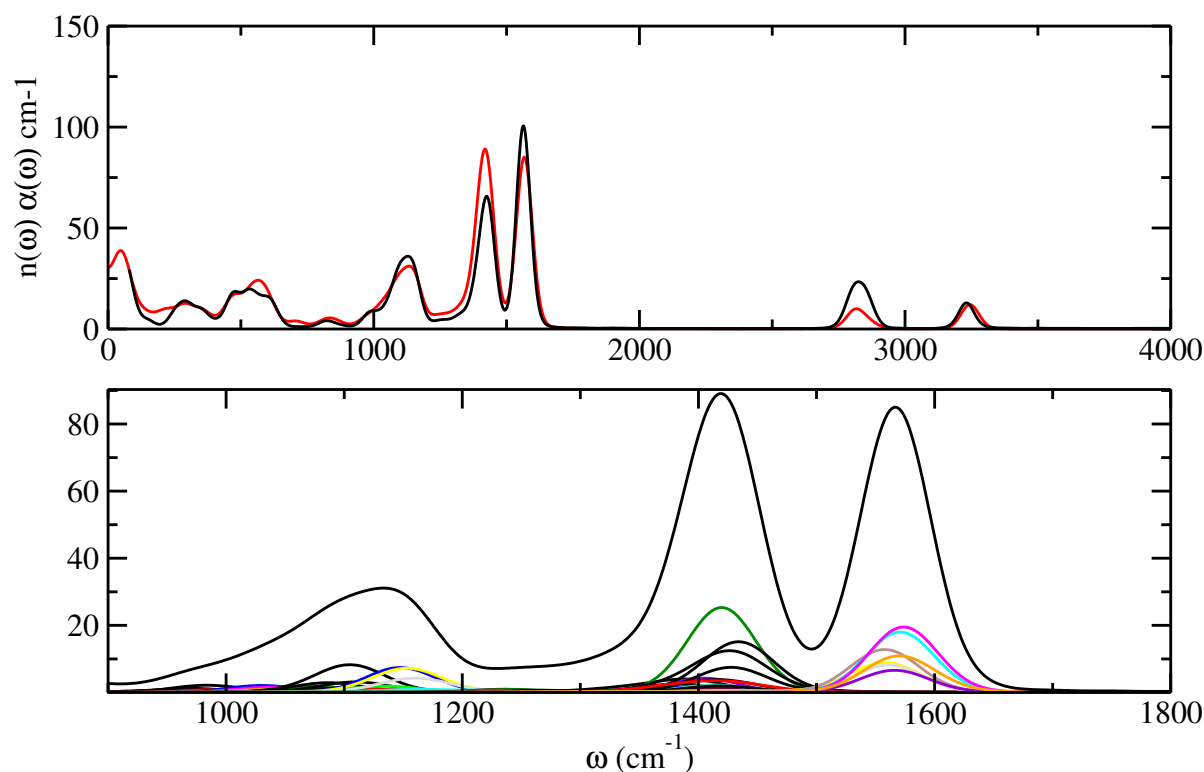


Figure 11. Gas phase octa-alanine at 20 K. Top: Comparison between the infrared spectrum calculated through the Fourier transform of the dipole correlation function (equation (8), black line) and the Fourier transform of the localized normal modes (equation (44), red line) making use of APT tensors (APT tensor calculated at the first configuration of the dynamics). Bottom: Decomposition of the infrared spectrum intensity into the different contributions arising from the Amide normal modes in the 1000-1800 cm^{-1} frequency domain: Dashed Black: total IR spectrum; Color curves: individual effective normal modes.